



Gagliardini, P., Ghysels, E., & Rubin, M. (2017). Indirect Inference Estimation of Mixed Frequency Stochastic Volatility State Space Models using MIDAS Regressions and ARCH Models. *Journal of Financial Econometrics*, 15(4), 509–560. [nbw013].
<https://doi.org/10.1093/jjfinec/nbw013>

Peer reviewed version

Link to published version (if available):
[10.1093/jjfinec/nbw013](https://doi.org/10.1093/jjfinec/nbw013)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Oxford University Press at <https://academic.oup.com/jfec/article/2977678/Indirects>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Indirect Inference Estimation of Mixed Frequency Stochastic Volatility State Space Models using MIDAS Regressions and ARCH Models*

P. Gagliardini[†] E. Ghysels[‡] M. Rubin[§]

First draft: October 2014

This version: November 14, 2016

*We thank the editors and referees for many helpful comments and suggestions which helped us greatly improve our paper. We also thank participants at the 2015 “8th Annual SoFiE Conference” in Aarhus, and the 2015 “11th World Congress of the Econometric Society”, in Montreal, for helpful comments.

[†]Università della Svizzera Italiana and Swiss Finance Institute (patrick.gagliardini@usi.ch).

[‡]University of North Carolina - Chapel Hill (eghysels@unc.edu).

[§]Università della Svizzera Italiana and Swiss Finance Institute (mirco.rubin@usi.ch).

Abstract

We examine the relationship between MIDAS regressions and the estimation of state space models applied to mixed frequency data. While in some cases the binding function is known, in general it is not, and therefore indirect inference is called for. The approach is appealing when we consider state space models which feature stochastic volatility, or other non-Gaussian and nonlinear settings where maximum likelihood methods require computationally demanding approximate filters. The stochastic volatility feature is particularly relevant when considering high frequency financial series. In addition, we propose a filtering scheme which relies on a combination of reprojection methods and nowcasting MIDAS regressions with ARCH models. We assess the efficiency of our indirect inference estimator for the stochastic volatility model by comparing it with the Maximum Likelihood (ML) estimator in Monte Carlo simulation experiments. The ML estimate is computed with a simulation-based Expectation-Maximization (EM) algorithm, in which the smoothing distribution required in the E step is obtained via a particle forward-filtering/backward-smoothing algorithm. Our Monte Carlo simulations show that the Indirect Inference procedure is very appealing, as its statistical accuracy is close to that of MLE but the former procedure has clear advantages in terms of computational efficiency. An application to forecasting quarterly GDP growth in the Euro area with monthly macroeconomic indicators illustrates the usefulness of our procedure in empirical analysis.

Keywords: Indirect inference, MIDAS regressions, State space model, Stochastic volatility, GDP forecasting.

1 Introduction

Econometric models that take into account the unbalanced nature of datasets have attracted substantial attention recently. Policy makers and practitioners alike need to assess in real-time the current state of the economy, with at best mixed frequency data at their disposal. For example, one of the key indicators of macroeconomic activity, the Gross Domestic Product (GDP), is released quarterly, while a range of leading and coincident indicators is timely available at a monthly or even higher frequency. Hence, we may want to construct a forecast of the current quarter GDP growth (a so called nowcast) based on the available higher frequency information.

Econometric models with mixed frequency data can be classified into two broad classes: (1) likelihood-based involving latent processes and (2) purely regression-based. The former category consists primarily of state space models, studied by Harvey and Pierse (1984), Harvey (1989), Zdrozny (1990), Bernanke, Gertler, and Watson (1997), Mariano and Murasawa (2003), Mitnik and Zdrozny (2005), Aruoba, Diebold, and Scotti (2009), Ghysels and Wright (2009), Kuzin, Marcellino, and Schumacher (2011), among others. The regression-based methods involve Mixed Data Sampling (MIDAS) regressions; see e.g. Ghysels, Santa-Clara, and Valkanov (2006), Andreou, Ghysels, and Kourtellis (2010). As one considers high frequency data, the issue of time-varying volatility becomes increasingly relevant. Dealing with stochastic volatility (SV) in state space models is doable but poses challenges both statistical and computational in nature. One possibility is to consider Bayesian approaches in this context, as done by Carriero, Clark, and Marcellino (2013) and Marcellino, Porqueddu, and Venditti (2016). However, when it comes to classical inference one typically relies on the Expectation-Maximization (EM) algorithm to compute numerically the ML estimate in a model with unobservable variables (Dempster, Laird, and Rubin (1977)). The likelihood function of the model involves a large-dimensional integral with respect to the latent factor paths as the latent factors appear in the conditional mean and volatility of the high frequency data series. This integral representation of the likelihood is impractical for the computation of the ML estimate.

If the objective is to estimate state space models with mixed frequency data - of which there are many examples - featuring stochastic volatility, using classical inference methods, is there perhaps a simpler way to do so? This is the contribution of our paper. We introduce indirect inference estimation procedures proposed by Gouriéroux, Monfort, and Renault

(1993), Smith (1993) and Gallant and Tauchen (1996), to estimate the models of interest using MIDAS regressions augmented with ARCH-type models as well as mixed frequency Vector Autoregressive (VAR) models (see e.g. Ghysels (2016)) as auxiliary models. Same frequency data settings are a special case of mixed frequency ones. The analysis in this paper is therefore also applicable to standard state space models. Moreover, the idea of estimating SV-type models using ARCH-type auxiliary models has a long history starting with Engle and Lee (1999) and Pastorello, Renault, and Touzi (2000). Our paper combines insights from the literature on SV models with those from the mixed frequency data literature.

It is worth noting that in some specific cases we know the binding function between the state space model and the implied MIDAS regression, as discussed in Bai, Ghysels, and Wright (2013). However, these cases are rather too simple to be practical, so that the use of indirect inference is a natural way to tackle the unknown binding function. The methods we propose are fairly easy to implement and involve auxiliary model-based estimators involving MIDAS regressions combined with ARCH specifications for the errors. In addition, we filter latent variables, given observables, using reprojection methods proposed by Gallant and Tauchen (1998).

We compare the two estimation methods, namely (1) Maximum Likelihood (ML) and (2) indirect inference, via Monte Carlo simulations. To implement the former method in the mixed frequency SV model, we consider a simulation-based estimator relying on the EM algorithm. The smoothing distribution required in the Expectation step is computed via a particle forward-filtering/backward-smoothing algorithm. We compare the two estimation methods on the basis of (a) statistical criteria - mean/bias/quantiles of sampling distributions, (b) filtering accuracy - both conditional mean and volatility and (c) computational time. Our results show that there are clear advantages in terms of computational time to the new indirect inference procedure put forward in this paper, while the losses in statistical efficiency compared to MLE are very limited. Even in the linear Gaussian case, we find our indirect inference methods remarkably accurate, when compared to the standard MLE based on the Kalman filter.

The paper is organized as follows. Section 2 introduces state space models with mixed frequency data and stochastic volatility. Section 3 defines our indirect inference estimator. This section covers the linear Gaussian state space model with mixed frequency data as a special case of the general specification, and discusses its relation with MIDAS regressions. This link yields useful insights to define the auxiliary model for indirect inference in the

general SV case. Section 3 also describes the estimation of the SV model with ML via a simulation-based EM algorithm. Section 4 discusses filtering via reprojection, followed by Section 5 which reports the results of an extensive Monte Carlo study. Section 6 presents an empirical application of our model to the problem of forecasting at short horizons Euro-area quarterly GDP growth using monthly macroeconomic indicators. The dataset is the same as the one considered in the empirical study of Marcellino, Porqueddu, and Venditti (2016). Section 7 concludes the paper.

2 State Space Models with Mixed Frequency Data and Stochastic Volatility

There is a burgeoning literature on nowcasting using either MIDAS regressions or state space models, see e.g. Mariano and Murasawa (2003), Nunes (2005), Giannone, Reichlin, and Small (2008), Aruoba, Diebold, and Scotti (2009), Marcellino and Schumacher (2010), Andreou, Ghysels, and Kourtellis (2013) and Banbura and Modugno (2014), among others. Recent surveys include Andreou, Ghysels, and Kourtellis (2011), Foroni and Marcellino (2013) and Banbura, Giannone, Modugno, and Reichlin (2013), where the latter paper has a stronger focus on more complex Kalman filter-based factor modeling techniques.

State space models have been widely used in econometrics as well as other scientific disciplines, in particular engineering where the Gaussian state space model and its Kalman filtering algorithm originated.¹ A key starting point is that observations are driven by some latent process. Moreover, it is also assumed that data are contaminated by measurement errors. To accommodate the mixed frequency sampling scheme, we adopt a time scale expressed in a form that easily represents such mixtures. We will focus on small values of m , the number of high frequency subperiods, such as for example $m = 3$ for monthly data sampled every quarter. We consider a dynamic model for the latent factors as follows:

Assumption 2.1 *Let (F) be a $n_f \times 1$ dimensional vector process satisfying*

$$F_{t+j/m} = \sum_{l=1}^p \Phi_l F_{t+(j-l)/m} + \eta_{t+j/m} \quad \forall t = 1, \dots, T, \quad j = 0, \dots, m-1, \quad (2.1)$$

¹The econometric literature on the topic is vast, see e.g. Harvey (1989), Hamilton (1994), among others.

where Φ_l are $n_f \times n_f$ matrices, the eigenvalues of the autoregressive matrix in the stacked $AR(1)$ representation lie inside the unit circle, and (η) is an i.i.d. zero mean Gaussian error process with diagonal covariance matrix $\Sigma_\eta = \text{diag}(\sigma_{i,\eta}^2, i = 1, \dots, n_f)$. Finally, the number of factors n_f , is assumed to be known.

We have two types of data: (1) time series sampled at a low frequency (LF) - every integer date t , and (2) time series sampled at high frequency (HF) - every $t + j/m$, with $j = 0, \dots, m - 1$. Bai, Ghysels, and Wright (2013) make two convenient simplifications which depart from generality. First, they assume that there is only one low-frequency process and call it y_t , and second, consider the combination of only two sampling frequencies. We will proceed with the same simplifications and also assume - for the sake of simplicity - that there is only one high-frequency series, denoted $x_{t+j/m}$. It is fairly easy to extend the methods proposed in this paper to cases involving multiple low and high frequency series - which we will not cover explicitly.

If the low-frequency process were observed at high frequency, it would relate to the factors as follows:

$$y_{t+j/m}^* = \gamma_1' F_{t+j/m} + u_{1,t+j/m} \quad \forall t, \quad j = 0, \dots, m - 1, \quad (2.2)$$

where y^* denotes the process which is not directly observed and γ_1 is a $n_f \times 1$ vector of factor loadings. The error process $u_{1,t+j/m}$ has an $AR(k)$ representation:

$$d_1(L^{1/m})u_{1,t+j/m} = \varepsilon_{1,t+j/m}, \quad d_1(L^{1/m}) \equiv 1 - d_{11}L^{1/m} - \dots - d_{k1}L^{k/m}, \quad (2.3)$$

where the lag operator $L^{1/m}$ applies to high-frequency data, i.e. $L^{1/m}u_t \equiv u_{t-1/m}$. The observed low-frequency process y relates to the process y^* via a linear aggregation scheme:

$$y_{t+j/m}^c = \Psi_j y_{t+(j-1)/m}^c + \lambda_j y_{t+j/m}^* \quad (2.4)$$

where y_t is equal to the cumulator variable y_t^c for integer t , and is not observed otherwise. The above scheme, also used by Harvey (1989) and Nunes (2005), covers both stock and flow aggregation. We get the case of a stock variable by setting $\Psi_j = 1 (j \neq 0, m, 2m, \dots)$ and $\lambda_j = 1 (j = 0, m, 2m, \dots)$, where $1(\cdot)$ denotes the indicator function. If we pick instead $\Psi_j = 1 (j \neq 1, m + 1, 2m + 1, \dots)$ and $\lambda_j = 1/m$ for all j , then we get a flow variable.

The high frequency process $x_{t+j/m}$ relates to the factors as follows:

$$x_{t+j/m} = \gamma_2' F_{t+j/m} + u_{2,t+j/m} \quad \forall t, \quad j = 0, \dots, m-1, \quad (2.5)$$

where γ_2 is a $n_f \times 1$ vector and:

$$d_2(L^{1/m})u_{2,t+j/m} = \varepsilon_{2,t+j/m}, \quad d_2(L^{1/m}) \equiv 1 - d_{12}L^{1/m} - \dots - d_{k2}L^{k/m}. \quad (2.6)$$

As usual in latent factor models, factor loadings γ_1 , γ_2 and the parameters of the factor dynamics are subject to identification restrictions.

The standard approach is to assume that the innovation processes (ε_k) are i.i.d. Gaussian with mean zero and variance $\sigma_{\varepsilon_k}^2$, for $k = 1, 2$. Indeed, the literature typically ignores the presence of time-varying volatility, yet the high frequency data often involve financial and other series which feature conditional heteroskedasticity. This means that the state space models are no longer Gaussian. There is a substantial literature on non-Gaussian state space models tailored for the analysis of financial returns data (see e.g. Ghysels, Harvey, and Renault (1996), Shephard (2005) and references therein). The type of models of interest to us are rather state space models with stochastic volatility in measurement equations. Hence, our analysis relates more directly to recent work by Clark (2011), Carriero, Clark, and Marcellino (2012), Carriero, Clark, and Marcellino (2013), or Marcellino, Porqueddu, and Venditti (2016).

We augment equations (2.5)-(2.6) for high frequency data with time-varying volatility:

$$\varepsilon_{2,t+j/m} \sim \mathcal{N}(0, h_{t+j/m}), \quad (2.7)$$

where the log volatility follows a Gaussian autoregressive process:

$$\ln h_{t+j/m} = c + \rho_{SV} \ln h_{t+(j-1)/m} + \xi_{t+j/m}, \quad \xi_{t+j/m} \sim i.i.d. \mathcal{N}(0, \nu_2^2), \quad (2.8)$$

and parameter ρ_{SV} is smaller than 1 in absolute value. We obtain a SV-type volatility specification without common factor structure.

While our analysis relates to recent work by Marcellino, Porqueddu, and Venditti (2016), among others, as noted before, there are also subtle but important differences. In their model the factor process features stochastic volatility. Instead, in equation (2.7) we assume

that the measurement error features stochastic volatility. When dealing with low frequency macroeconomic series exposed to factors, we think it is more appropriate to assume that those factors do not feature volatility clustering, while the high frequency series are conditionally heteroskedastic. Ideally one could consider models where SV is featured in both the observation and state equations. This is of course a model choice decision. SV or ARCH features, while prominently present in high frequency series, are diminished in importance when it comes to low frequency phenomena. Temporal aggregation is one argument. For our analysis, the factor tracks low frequency data, and retrieves that information from high frequency data as well - albeit contaminated with noise that indeed features volatility clustering. We leave this as a topic for future research. Assumptions 2.1 and 2.2 (below) define the parametric models of interest in this paper. We denote by θ the vector of unknown parameters in these models.

Assumption 2.2 *The observable processes (y) and (x) are such that:*

$$\begin{aligned}
y_{t+j/m}^* &= \gamma_1' F_{t+j/m} + u_{1,t+j/m}, \\
d_1(L^{1/m})u_{1,t+j/m} &= \varepsilon_{1,t+j/m}, \quad d_1(L^{1/m}) \equiv 1 - d_{11}L^{1/m} - \dots - d_{k1}L^{k/m}, \\
y_{t+j/m}^c &= \Psi_j y_{t+(j-1)/m}^c + \lambda_j y_{t+j/m}^*, \\
y_t &= y_t^c, \\
x_{t+j/m} &= \gamma_2' F_{t+j/m} + u_{2,t+j/m}, \\
d_2(L^{1/m})u_{2,t+j/m} &= h_{t+j/m}^{1/2} \varepsilon_{2,t+j/m}, \quad d_2(L^{1/m}) \equiv 1 - d_{12}L^{1/m} - \dots - d_{k2}L^{k/m}, \\
\ln h_{t+j/m} &= c + \rho_{SV} \ln h_{t+(j-1)/m} + \xi_{t+j/m}, \quad \forall t, \quad j = 0, \dots, m-1
\end{aligned}$$

where $|\rho_{SV}| < 1$, and (ε_1) , (ε_2) , (ξ) are mutually independent i.i.d. Gaussian processes, with distributions $\mathcal{N}(0, \sigma_{\varepsilon_1}^2)$, $\mathcal{N}(0, 1)$, $\mathcal{N}(0, \nu_2^2)$ respectively, and independent of process (η) .

3 Indirect Inference Estimation

Estimating via Maximum Likelihood (ML) the mixed frequency models with SV presented in the previous section is rather involved. Indeed, the likelihood function involves a large-dimensional integral with respect to the latent factors path. This integral representation of the likelihood is impractical for computation of the ML estimate, and numerical filtering techniques are necessary.

In this section we introduce indirect inference estimation methods - proposed by Gouriéroux, Monfort, and Renault (1993), Smith (1993) and Gallant and Tauchen (1996) - to estimate the mixed frequency SV models. Indirect inference can be used to estimate virtually any model from which it is possible to simulate data. This obviously includes state space models. Indirect inference estimation in fact involves two types of models - a model of interest already specified in the previous section - and an auxiliary model which is easy to estimate. Both models are linked - in terms of parameter spaces - by a binding function.

3.1 Linear Setting with Known Binding Function

To explain our estimation approach it is worth starting with a setting where the binding function is known. This setting is provided by a linear state space model with Gaussian errors. This model is a special case of the general specification in Assumptions 2.1 and 2.2 when there is no SV.² In this linear state space model, the Kalman filter can be applied for prediction and filtering. Bai, Ghysels, and Wright (2013) show that for a model with a single latent factor ($n_f = 1$) having a AR(1) dynamics and persistence parameter ρ , and $m = 3$ as for instance for a monthly/quarterly mixture of data, one obtains (see Appendix A.2 for details):

$$E[y_{t+h}|I_t^M] = \rho^{3h}\kappa_{3,1} \sum_{j=0}^{\infty} \vartheta^j y_{t-j} + \rho^{3h} \sum_{j=0}^{\infty} \vartheta^j x(\theta_x)_{t-j} \quad (3.1)$$

where I_t^M denotes the information in the available low and high frequency data up to time t , $\vartheta = [(\rho - \rho\kappa_1)(\rho - \rho\kappa_2)(\rho - \rho\kappa_3)]$, and κ_i , $\kappa_{3,i}$ are steady state Kalman gain parameters. Moreover, one has:

$$x(\theta_x)_t \equiv [\kappa_{3,2} + (\rho - \rho\kappa_3)\kappa_2 L^{1/3} + (\rho - \rho\kappa_3)(\rho - \rho\kappa_2)\kappa_1 L^{2/3}]x_t \quad (3.2)$$

which is a parameter-driven low-frequency process composed of high-frequency data aggregated at the quarterly level.

The above equation relates to the multiplicative MIDAS regression models considered by Chen and Ghysels (2010) and Andreou, Ghysels, and Kourtellis (2013). In particular

²It corresponds to the parameter constraints $\nu_2 = 0$ and $\rho_{SV} = 1$.

consider the following ADL-MIDAS regression:

$$y_{t+h} = \beta_y \sum_{j=0}^{K_y} w_j(\theta_y) y_{t-j} + \beta_x \sum_{j=0}^{K_x} w_j(\theta_x^1)^j x(\theta_x^2)_{t-j} + \varepsilon_{t+h} \quad (3.3)$$

where $w_j(\theta_y)$, $w_j(\theta_x^1)$ follow an exponential Almon scheme and

$$x(\theta_x^2)_{t-j} \equiv \sum_{k=0}^{m-1} w_k(\theta_x^2) L^{k/m} x_{t-k/m}$$

also follows an exponential Almon scheme.³ Provided that $\rho > 0$, equation (3.1) is a special case of this model with $K_y = K_x = \infty$, $w_j(\theta_y) \propto \exp(\log(\vartheta)j)$, $w_j(\theta_x^1) \propto \exp(\log(\vartheta)j)$ and $w_k(\theta_x^2) \propto \exp(\theta_{x,1}^2 k + \theta_{x,2}^2 k^2)$ where $\theta_{x,1}^2$ and $\theta_{x,2}^2$ are parameters that solve the equations:

$$\begin{aligned} \log\{(\rho - \rho\kappa_3)\kappa_2/\kappa_{3,2}\} &= \theta_{x,1}^2 + \theta_{x,2}^2, \\ \log\{(\rho - \rho\kappa_3)(\rho - \rho\kappa_2)\kappa_1/\kappa_{3,2}\} &= 2\theta_{x,1}^2 + 4\theta_{x,2}^2. \end{aligned} \quad (3.4)$$

Equations (3.3) and (3.4) implicitly define a binding function between the parameters of the state space model and those of the MIDAS regression. Note, however, that the mapping under-identifies the parameters of the state space model if we rely on a standard multiplicative MIDAS regression scheme. Moreover, the mapping is only valid for a single factor state space model with i.i.d. measurement errors. What do we do for multi-factor models or single factor models with autoregressive errors? Bai, Ghysels, and Wright (2013) show that MIDAS regressions still provide very accurate approximations, although there is no exact (underidentified) mapping.

3.2 Auxiliary Models: U-MIDAS and ARCH

A departure from the setup in Bai, Ghysels, and Wright (2013) is that we replace equation (3.3) with a U-MIDAS - meaning unrestricted MIDAS - specification suggested by Foroni,

³The constructed low-frequency regressor is estimated jointly with the other (MIDAS) regression parameters. Hence, one can view $x(\theta_x^2)_{t-j}$ as the best *aggregator* that yields the best *prediction*. This ADL-MIDAS regression involves more parameters than the usual specification involving only one polynomial.

Marcellino, and Schumacher (2013), namely:

$$y_{t+1} = \bar{\beta}_0 + \sum_{k=0}^{\tilde{K}_y} \beta_k y_{t-k} + \sum_{j=1}^{m(\tilde{K}_x+1)-1} \gamma_j x_{t+1-j/m} + \varepsilon_{t+1}. \quad (3.5)$$

Note that we estimate $\tilde{K}_y + m(\tilde{K}_x + 1)$ parameters (not including intercept and residual variance). When m is small, as shown by Foroni, Marcellino, and Schumacher (2013), we are able to estimate these parameters with reasonable precision using sample sizes typically encountered in economic applications. One attractive feature of U-MIDAS misspecification is the fact that estimation is numerically straightforward, as it can be performed by OLS.

Suppose we collect all the parameters of the U-MIDAS regression into the vector $\phi \in \Phi$. Assuming $\dim(\theta) \leq \dim(\phi) \equiv \tilde{K}_y + m(\tilde{K}_x + 1) + 2$ we may be able to identify and estimate the parameters via indirect inference.⁴

Since the models of interest feature SV, we can consider as auxiliary models the following U-MIDAS regressions augmented with ARCH errors:

$$\begin{aligned} y_{t+1} &= \bar{\beta}_0 + \sum_{k=0}^{\tilde{K}_y} \beta_k y_{t-k} + \sum_{j=1}^{m(\tilde{K}_x+1)-1} \gamma_j x_{t+1-j/m} + \varepsilon_{t+1}, & \varepsilon_{t+1} &\sim \mathcal{N}(0, \sigma_{t+1}^2) \\ \sigma_t^2 &= \omega + \sum_{k=1}^p \alpha_k \varepsilon_{t-k}^2 \end{aligned} \quad (3.6)$$

which has the advantage of being simple to implement as it only involves a linear regression specification with ARCH(p) errors. The idea for this auxiliary model is that heteroskedasticity in the high frequency data affects the residuals of the reduced form MIDAS regressions. Obviously, the ARCH model in the above equation is only estimated at low frequency, and therefore the ARCH effects may not be particularly strong.

3.3 Auxiliary Models: Mixed Frequency VAR and ARCH

The auxiliary U-MIDAS regressions considered in the previous subsection do not fully exploit all features of the data since the link between latent factors and high frequency data is not being taken into account. In this subsection we remedy to this shortcoming by considering

⁴Note that we added a constant and residual variance in the MIDAS regressions parameter count.

mixed frequency VAR models. It is worth noting from the start that there might be some confusion about the characterization of mixed frequency VAR models. The analysis below serves two purposes: (1) it generalizes the U-MIDAS setup discussed so far and (2) it enables us to consider a suitable approach for state space models with stochastic volatility.

A number of authors, including Zdrozny (1988), Zdrozny (1990) and more recently Kuzin, Marcellino, and Schumacher (2011), Schorfheide and Song (2013), among others, start from a *latent* high frequency VAR process, namely:

$$\begin{pmatrix} y_{t+(j+1)/m}^* \\ x_{t+(j+1)/m} \end{pmatrix} = C_0 + \sum_{k=1}^{k^{max}} C_k \begin{pmatrix} y_{t+(j+1-k)/m}^* \\ x_{t+(j+1-k)/m} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t+(j+1)/m}^{y*} \\ \varepsilon_{t+(j+1)/m}^x \end{pmatrix} \quad (3.7)$$

where $y_{t+j/m}^*$ is defined in equation (2.2). The above latent VAR model is related to observables via a measurement equation and therefore cast in state space framework with missing observations.

State space models are, using the terminology of Cox (1981), parameter-driven models whereas VAR models are, using again the same terminology, observation-driven models as they are formulated exclusively in terms of observable data. Ghysels (2016) introduces a class of observation-driven mixed frequency VAR models which provides an alternative to commonly used state space models involving latent processes. In addition, the mixed frequency VAR model is a multivariate extension of MIDAS regressions.

The mixed frequency VAR considered by Ghysels (2016), tailored towards the current application, can be written as follows:

$$\begin{pmatrix} x_{t+1} \\ \vdots \\ x_{t+1+(m-1)/m} \\ y_{t+1} \end{pmatrix} = \tilde{C}_0 + \sum_{k=1}^{\tilde{K}^{max}} \tilde{C}_k \begin{pmatrix} x_{t+1-k} \\ \vdots \\ x_{t+1-k+(m-1)/m} \\ y_{t+1-k} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t+1}^1 \\ \vdots \\ \varepsilon_{t+1}^m \\ \varepsilon_{t+1}^y \end{pmatrix}. \quad (3.8)$$

Hence, it involves a VAR of dimension $m + 1$ (with single high and low frequency series) where the high and low frequency data for low frequency period (quarter, say) t are stacked into a vector whose dynamics is described by a linear multivariate autoregressive structure. Note, that elements of the matrices \tilde{C}_k now describe within-period (intra-quarterly) time series dependencies.⁵ The stacking implies that, if we read across a particular row of the

⁵Most notably Granger causal patterns as discussed in Ghysels, Hill, and Motegi (2014), Ghysels, Hill,

mixed frequency VAR, we have high frequency processes predicted by past high and low frequency series and vice versa.

The unrestricted VAR model in equation (3.8) includes $(m+1) + \tilde{K}_{max}(m+1)^2 + m(m+1)/2$ parameters which can be estimated by OLS. Ghysels (2016) proposes a parsimonious parametrization which can be estimated by Maximum Likelihood, at the expense of a higher computational cost, as the likelihood has to be maximized numerically either using classical or Bayesian techniques. Therefore, this restricted VAR is not suitable as auxiliary model, because of the heavy computational cost.

A parsimonious auxiliary model, which ensures computational speed for indirect inference estimation, can be obtained by considering an U-MIDAS regression model for the low frequency data, and an AR model for the high frequency ones. In fact, both models can be easily estimated by OLS. Therefore, the following model will be used as the auxiliary model in our Monte Carlo simulation exercise for DGPs without SV:

$$\begin{cases} y_{t+1} &= \bar{\beta}_0 + \sum_{k=0}^{\tilde{K}_y} \beta_k y_{t-k} + \sum_{j=1}^{m(\tilde{K}_x+1)-1} \gamma_j x_{t+1-j/m} + \zeta_{t+1}^y \\ x_{t+(j+1)/m} &= c_0 + \sum_{k=1}^{m(\tilde{K}_x+1)} c_k x_{t+(j+1-k)/m} + \zeta_{t+(j+1)/m}^x \end{cases} \quad (3.9)$$

The first equation of this auxiliary model corresponds to the U-MIDAS specification in equation (3.5), while the second equation is an AR of order $m(\tilde{K}_x + 1)$ specified on the high frequency (HF) data only. Stacking the low and high frequency data into a vector, we note that the set of equations in (3.9) amounts to a mixed frequency VAR model as defined in (Ghysels 2016), where we impose the restriction that low frequency data do not Granger cause the high frequency series.⁶ Specifically, the first equation in model (3.9) corresponds to the last equation in the mixed frequency VAR model (3.8), while the second equation in model (3.9) corresponds to the first equation in the mixed frequency VAR model (3.8), where the high frequency variables do not depend explicitly on the lagged low frequency ones.

Model (3.9) can be estimated by OLS, and the correlation between the innovations ζ_{t+1}^y and $\zeta_{t+(j+1)/m}^x$, which can only be computed at low frequency, could be included as an auxiliary parameter to estimate, or can be set to zero.

and Motegi (2016), Götz and Hecq (2014a) and Götz and Hecq (2014b).

⁶Although this is a simplifying hypothesis which might be relaxed, our model produced remarkably good results in the extensive Monte Carlo Simulation exercises described in Section 5.

In order to handle the DGP with SV, we can add ARCH-type augmentations to the auxiliary models. In particular, the complete auxiliary model used in the Monte Carlo simulation for DGPs with SV is:

$$\left\{ \begin{array}{lcl} y_{t+1} & = & \bar{\beta}_0 + \sum_{k=0}^{\tilde{K}_y} \beta_k y_{t-k} + \sum_{j=0}^{m(\tilde{K}_x+1)-1} \gamma_j x_{t+1-j/m} + \zeta_{t+1}^y \\ x_{t+(j+1)/m} & = & c_0 + \sum_{k=1}^{m(\tilde{K}_x+1)} c_k x_{t+(j+1-k)/m} + \zeta_{t+(j+1)/m}^x, \quad \zeta_{t+(j+1)/m}^x \sim \mathcal{N}(0, \sigma_{t+(j+1)/m}^x) \\ \sigma_{t+(j+1)/m}^x & = & \omega + \sum_{k=1}^p \alpha_k (\zeta_{t+(j+1-k)/m}^x)^2 \end{array} \right. \quad (3.10)$$

where the errors ζ^x and ζ^y can be correlated.

3.4 Estimation via Indirect Inference

The parameter vectors for the auxiliary model will be denoted respectively ϕ^{Mi} for the U-MIDAS specification appearing in equation (3.6), and ϕ^V for the mixed frequency VAR model in equation (3.8) in general - and more specifically in equation (3.10). Given a sample of size Tm we obtain OLS estimates $\hat{\phi}_{Tm}^{Mi}$ and $\hat{\phi}_{Tm}^V$.

We simulate mixed frequency data with the state space model in Assumptions 2.1 and 2.2, given a particular structural parameter value θ , by drawing S independent samples of size

$T^S m$ from the model:

$$\begin{aligned}
F_{s,t+j/m}(\theta) &= \sum_{l=1}^p \Phi_l(\theta) F_{s,t+(j-l)/m}(\theta) + \eta_{s,t+j/m}(\theta) \\
\ln h_{s,t+j/m}(\theta) &= c(\theta) + \rho_{SV}(\theta) \ln h_{s,t+(j-1)/m}(\theta) + \xi_{s,t+j/m}(\theta) \\
y_{s,t+j/m}^*(\theta) &= \gamma_1(\theta)' F_{t+j/m}(\theta) + u_{s,1,t+j/m}(\theta) \\
d_1(L^{1/m}, \theta) u_{s,1,t+j/m}(\theta) &= \varepsilon_{s,1,t+j/m}(\theta) \\
y_{s,t+j/m}^c(\theta) &= \Psi_j y_{s,t+(j-1)/m}^c(\theta) + \lambda_j y_{s,t+j/m}^*(\theta) \\
y_{s,t}(\theta) &= y_{s,t}^c(\theta) \\
x_{s,t+j/m}(\theta) &= \gamma_2(\theta)' F_{s,t+j/m}(\theta) + u_{s,2,t+j/m}(\theta) \\
d_2(L^{1/m}, \theta) u_{s,2,t+j/m}(\theta) &= h_{s,t+j/m}(\theta)^{1/2} \varepsilon_{s,2,t+j/m}(\theta) \\
\forall t = 1, \dots, T^S, \quad j = 0, \dots, m-1, \quad s = 1, \dots, S, \quad (3.11)
\end{aligned}$$

where innovation processes $\eta_s(\theta)$, $\varepsilon_{s,1}(\theta)$, $\varepsilon_{s,2}(\theta)$ and $\xi_s(\theta)$ are independent i.i.d. processes with Gaussian distributions $\mathcal{N}(0, \Sigma_\eta(\theta))$, $\mathcal{N}(0, \sigma_{\varepsilon_1}^2(\theta))$, $\mathcal{N}(0, 1)$, $\mathcal{N}(0, \nu_2^2(\theta))$. Given the S simulated samples, we compute the following estimators:

- The Indirect Inference (II) estimator of Gouriéroux, Monfort, and Renault (1993) and Smith (1993), using the U-MIDAS auxiliary model, denoted by $\hat{\theta}_{TmS}^{IIMi}$;
- The II estimator of Gouriéroux, Monfort, and Renault (1993) and Smith (1993), using the mixed frequency VAR auxiliary model, denoted by $\hat{\theta}_{TmS}^{IV}$.

The II estimators for auxiliary models Mi and V are obtained via:

$$\hat{\theta}_{TmS}^{IIMi} = \arg \min_{\theta} \left(\hat{\phi}_{Tm}^{Mi} - \frac{1}{S} \sum_s \hat{\phi}_{Tm,s}^{Mi}(\theta) \right)' \Omega^{Mi} \left(\hat{\phi}_{Tm}^{Mi} - \frac{1}{S} \sum_s \hat{\phi}_{Tm,s}^{Mi}(\theta) \right) \quad (3.12)$$

and:

$$\hat{\theta}_{TmS}^{IV} = \arg \min_{\theta} \left(\hat{\phi}_{Tm}^V - \frac{1}{S} \sum_s \hat{\phi}_{Tm,s}^V(\theta) \right)' \Omega^V \left(\hat{\phi}_{Tm}^V - \frac{1}{S} \sum_s \hat{\phi}_{Tm,s}^V(\theta) \right) \quad (3.13)$$

respectively, with $\hat{\phi}_{Tm,s}^{Mi}(\theta)$ and $\hat{\phi}_{Tm,s}^V(\theta)$ being the U-MIDAS and VAR auxiliary model parameter estimates for generated sample s and structural parameter value θ , and Ω^{Mi} and Ω^V being (optimal) weighting matrices.

Assumptions 2.1-2.2 and standard regularity conditions (see e.g. Gouriéroux and Monfort (1997)) imply that the indirect inference estimators are consistent and asymptotically normal as T and $S \rightarrow \infty$:

$$\sqrt{Tm}(\hat{\theta}_{TmS}^{EST} - \theta_0) \rightarrow_d \mathcal{N}(0, V^{EST}) \quad (3.14)$$

for $EST \equiv IIMi$ and IIV respectively, where θ_0 denotes the true value of the structural parameter.

3.5 EM algorithm for mixed frequency SV model

We assess the efficiency of our indirect inference estimators for the SV model with mixed frequency data by comparing their performances with that of the Maximum Likelihood (ML) estimator in a Monte Carlo experiment (see Section 5). Due to the latent factor processes (F) and (h) in the dynamics of the data, the likelihood function of the model involves a large-dimensional integral with respect to the latent factors path. This integral representation of the likelihood is impractical for computation of the ML estimate. We consider instead a simulation-based estimator relying on the Expectation Maximization (EM) algorithm. The smoothing distribution required in the Expectation step is computed via a particle forward-filtering/backward-smoothing algorithm.⁷ In this section we describe the main steps of the procedure, and refer to Appendix B for the detailed definition of the estimation algorithm.

The Expectation-Maximization (EM) algorithm is an iterative procedure to compute numerically the ML estimate in a model with unobservable variables (Dempster, Laird, and Rubin (1977)). Let $Y_t = (y_t, x_{t-j/m}, j = 0, 1, \dots, m-1)'$ be the vector of stacked observable variables (measurements) and $f_t = (F_{t-j/m}, h_{t-j/m}, j = 0, 1, \dots, m-1)'$ the Markov vector of stacked latent factors, for $t = 1, \dots, T$. The EM algorithm relies on the complete-observation log-likelihood function, that is the log of the joint density of observable and unobservable variables in the structural model:

$$\begin{aligned} \mathcal{L}^*(\theta) &= \log \ell(\underline{Y}_T, \underline{f}_T; \theta) \\ &= \sum_{t=1}^T \log h(Y_t | Y_{t-1}, \underline{f}_t; \theta) + \sum_{t=1}^T \log g(f_t | f_{t-1}; \theta), \end{aligned}$$

⁷Other approaches have been proposed in the literature to implement the MLE in nonlinear state space models with SV and could be adapted to our mixed frequency framework, for instance the Monte Carlo ML approach in Sandmann and Koopman (1998).

where \underline{Y}_T denotes the history of Y_t up to T , and similarly for \underline{f}_T and \underline{f}_t . Here, h is the measurement density and g is the transition density in the state space representation (see Appendix B.2 for the expression of $\mathcal{L}^*(\theta)$ in the mixed-frequency SV model). Let $\hat{\theta}_{Tm}^{EM,(i)}$ be the estimate of parameter θ at iteration i of the EM algorithm. The update $i \rightarrow i + 1$ consists of two steps:

1. *Expectation (E) step.* Compute function $Q(\theta|\tilde{\theta})$, with $\tilde{\theta} = \hat{\theta}_{Tm}^{EM,(i)}$, where:

$$Q(\theta|\tilde{\theta}) = E_{\tilde{\theta}}[\mathcal{L}^*(\theta)|\underline{Y}_T]$$

and $E_{\tilde{\theta}}[\cdot|\underline{Y}_T]$ denotes the expectation w.r.t. the conditional distribution of \underline{f}_T given \underline{Y}_T for parameter value $\tilde{\theta}$.

2. *Maximization (M) step.* Compute the estimate for iteration $i + 1$ as:

$$\hat{\theta}_{Tm}^{EM,(i+1)} := \arg \max_{\theta} Q(\theta|\hat{\theta}_{Tm}^{EM,(i)}).$$

The iteration is performed until a criterion for numerical convergence of the estimate is met, and $\hat{\theta}_{Tm}^{ML} = \hat{\theta}_{Tm}^{EM,(\infty)}$. The details for the E-step and the M-step in the mixed frequency SV model are provided in Appendix B.3.

The E-step in the EM algorithm requires the smoothing distribution of the unobservable factor path for given parameter value $\tilde{\theta}$ to compute the conditional expectation $E_{\tilde{\theta}}[\cdot|\underline{Y}_T]$. This smoothing distribution cannot be characterized analytically for a nonlinear state space specification as the mixed frequency SV model. We approximate the smoothing distribution via a large sample of draws from it, called particles. The smoothing algorithm we adopt uses a sample of particles from the filtering distribution as an input. Specifically, for the E-step of the i -th iteration in the EM algorithm, we generate samples $f_{t+j/m}^{s,(i)} = (F_{t+j/m}^{s,(i)}, h_{t+j/m}^{s,(i)})'$, $s = 1, \dots, S$, from the filtering distribution of the latent factors at each date $t + j/m$, for parameter value $\hat{\theta}_{Tm}^{EM,(i)}$. For this task we use a sequential algorithm based on the auxiliary particle filter method running from the first sample date to the last sample date. We refer to Pitt and Shephard (1999) for the auxiliary particle filter; see also e.g. Douc, Moulines, and Olsson (2009), Carvalho, Johannes, Lopes, and Polson (2010), Doucet (2010), Lopes and Tsay (2011), Creal (2012), Kantas, Doucet, Singh, Maciejowski, and Chopin (2015) for recent developments and applications. The algorithm is described in detail in Section B.4.2. Then, we use a backward algorithm to generate sample paths $(\tilde{f}_{t+j/m}^{s,(i)}, \forall t, j)$, $s = 1, \dots, S$,

from the smoothing distribution; see e.g. Kim and Stoffer (2008) and Godsill, Doucet, and West (2004). Appendix B.4.3 provides the detailed simulation procedure. The sample paths $(\tilde{f}_{t+j/m}^{s,(i)}, \forall t, j)$, $s = 1, \dots, S$, are approximate draws from the distribution of $(f_{t+j/m}, \forall t, j)$ given \underline{Y}_T for parameter value $\hat{\theta}_{Tm}^{EM,(i)}$ when the number of particles S is large. We use averages across these sample paths to approximate the conditional expectation $E_{\tilde{\theta}}[\cdot | \underline{Y}_T]$ for $\tilde{\theta} = \hat{\theta}_{Tm}^{EM,(i)}$.

4 Filtering via reprojection and nowcasting

State space models do not only involve parameter estimation but also filtering of the latent states, for which the Kalman filter is the standard scheme in the linear Gaussian case. In this section we present alternative methods which easily extend to, say, the non-Gaussian case involving stochastic volatility. Our approach relies on the reprojection method of Gallant and Tauchen (1998) to produce filtering estimates of the latent factors.

The procedure is fairly simple to implement. Let $\hat{\theta}_{TmS}^{EST}$ be the parameter estimate obtained by one of the Indirect Inference estimators introduced in Section 3.4. We start again with simulating a long sample of size $T^{reproj}m$, say, from the model of interest as in equation (3.11), using parameter value $\theta = \hat{\theta}_{TmS}^{EST}$. Then, the simulated sample is used to estimate a specification for the conditional expectation of the latent factors given the observable data. Finally, the estimated specification for this conditional expectation is applied to the original sample of observations y , x , and used as a filter.

To develop further insight in the methodology, we start with the Gaussian case (the model in Assumptions 2.1 and 2.2 without SV, i.e. with $\rho_{SV} = 1$ and $\nu_2 = 0$). Next, we discuss the filtering algorithm for the non-Gaussian case with stochastic volatility.

In a Gaussian linear state space model, the conditional expectation of the latent factor given the measurements is linear (see equation (A.15) in Appendix A). In our mixed frequency setting, in analogy to equation (3.5), this remark suggests estimating a U-MIDAS regression

on the simulated sample:

$$\begin{aligned}
F_{t+j/m}(\hat{\theta}_{TmS}^{EST}) &= b_0(\hat{\theta}_{TmS}^{EST}) + \sum_{k=0}^{\tilde{K}_y} b_k(\hat{\theta}_{TmS}^{EST}) y_{t-k}(\hat{\theta}_{TmS}^{EST}) \\
&\quad + \sum_{k=0}^{m\tilde{K}_x} c_k(\hat{\theta}_{TmS}^{EST}) x_{t+(j-k)/m}(\hat{\theta}_{TmS}^{EST}) + \epsilon_{t+j/m}(\hat{\theta}_{TmS}^{EST}) \\
t &= 1, 2, \dots, T^{reproj}, \quad j = 0, 1, \dots, m-1,
\end{aligned} \tag{4.1}$$

which amounts to regressing latent factors onto observables. Note that the observables have a nowcasting feature, i.e. contemporaneous period $t + j/m$ high frequency data is used. Once we have the parameters of the above regression, we can apply the scheme to observed data y and x and therefore use it as a filter. We denote by $\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{TmS}^{EST})$ the reprojection factor values.

Likewise, the mixed frequency VAR framework of Ghysels (2016) could be modified to perform the task as filter, namely we run the system of regressions:

$$\begin{aligned}
\bar{C}(\hat{\theta}_{TmS}^{EST}) \begin{pmatrix} F_{t+1}(\hat{\theta}_{TmS}^{EST}) \\ \vdots \\ F_{t+1+(m-1)/m}(\hat{\theta}_{TmS}^{EST}) \\ y_{t+1}(\hat{\theta}_{TmS}^{EST}) \end{pmatrix} &= \tilde{C}_0(\hat{\theta}_{TmS}^{EST}) + \sum_{k=1}^{\tilde{K}_{max}} \tilde{C}_k(\hat{\theta}_{TmS}^{EST}) \begin{pmatrix} x_{t+1-k}(\hat{\theta}_{TmS}^{EST}) \\ \vdots \\ x_{t+1-k+(m-1)/m}(\hat{\theta}_{TmS}^{EST}) \\ y_{t+1-k}(\hat{\theta}_{TmS}^{EST}) \end{pmatrix} \\
&\quad + \begin{pmatrix} \epsilon_{t+1}^1(\hat{\theta}_{TmS}^{EST}) \\ \vdots \\ \epsilon_{t+1}^m(\hat{\theta}_{TmS}^{EST}) \\ \epsilon_{t+1}^y(\hat{\theta}_{TmS}^{EST}) \end{pmatrix}
\end{aligned} \tag{4.2}$$

where $\bar{C}(\hat{\theta}_{TmS}^{EST})$ is a lower triangular matrix to accommodate nowcasting - see Ghysels (2016) for further details. Here again, once we estimate the system of equations over a long simulated sample, we can treat the resulting estimates as weights for a filtering scheme.

In the nonlinear state space model with stochastic volatility, the conditional expectation of the latent factors given the current and past values of the observable variables is no more linear in the conditioning variables. Therefore, in such framework the regressions in (4.1) and (4.2) do not provide exact filters (up to a truncation of the number of lags). However, we can interpret these regressions as numerically feasible linear approximations of the unknown

exact filter for the latent factor F in the conditional mean. A second-order approximation is obtained by including quadratic terms in low and high frequency observations. Similar approximate filters can be developed for the stochastic volatility factor h . In this case, the filter can be based on squared measurement errors. For instance, in a model without AR effects in the measurement errors at high frequency (to simplify), we can run the regression:

$$h_{t+j/m}(\hat{\theta}_{TmS}^{EST}) = \bar{d}_0 + \sum_{k=0}^{\bar{K}_u} \bar{d}_k (u_{2,t+(j-k)/m}(\hat{\theta}_{TmS}^{EST}))^2 + \epsilon_{t+j/m}^u(\hat{\theta}_{TmS}^{EST}), \quad (4.3)$$

possibly including also higher-order terms.

In the linear Gaussian case, we can make direct comparisons of the filters based on reprojection with the Kalman filter in order to gauge the reliability of the proposed method. In the non-Gaussian case with stochastic volatility, a benchmark for comparison is obtained by first estimating the model by Monte Carlo EM as described in Section 3.5, and then compute the filtered factor value $\hat{f}_{t+j/m|t+j/m}(\hat{\theta}_{Tm}^{ML}) = [\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{Tm}^{ML}), \hat{h}_{t+j/m|t+j/m}(\hat{\theta}_{Tm}^{ML})]'$, say, by averaging the particles $f_{t+j/m}^s$, with $s = 1, \dots, S$, from the filtering distribution for parameter value $\hat{\theta}_{Tm}^{ML}$. We perform these comparisons in the Monte Carlo simulations presented in Section 5. There, we keep the reprojections quite simple in fact, namely we implement the filter for F in equation (4.1) in both the Gaussian and stochastic volatility settings, and we use a filter for volatility factor h based on squared residuals such as (4.3) in the latter setting. These filters could be improved upon by considering the setup in equation (4.2), and adding higher-order terms in the SV case.

Finally, let us recall that for all our DGPs we made the simplifying choice of not having a new equation for the volatility dynamics of the latent factor F_t . Nevertheless, the parameters for the dynamics of the SV of F_t - let us denote this new factor by g_t - could be estimated using the same auxiliary model as in (3.10), with the possibility of augmenting the system of equations with an ARCH-type equation for the low frequency innovation term ζ^y , in order to increase the information content on the auxiliary parameters for the new SV process. Analogously, the latent factor F_t could be filtered adapting the reprojection methodology of this section to include quadratic - and higher - order terms in low and high frequency observations in equations (4.1) and (4.2). Finally, g_t could also be estimated by reprojection of its simulated values on η_t (i.e. the simulated innovation terms of the latent factor F_t) and their higher order terms, together with other observable variables as well.

5 Monte Carlo Simulations

We conduct a Monte Carlo simulation to appraise the small and large sample properties of the indirect inference procedures proposed in earlier sections. A first subsection covers the design of the simulations. A second subsection covers the Gaussian state space model where the Kalman filter and maximum likelihood are the natural benchmarks. In a final subsection we consider non-Gaussian cases with stochastic volatility, where we compare our indirect inference procedure with a simulation-based EM algorithm.

5.1 Design

We consider three designs for the MC experiments. In all of them we have $m = 3$, corresponding to - for instance - a mixture of monthly and quarterly data, and stock sampling of the low frequency variable. In the first MC design, we consider a linear Gaussian state space model. The DGP has a single Gaussian AR(1) latent factor process ($n_f = 1$), and Gaussian AR(1) measurement errors for both the high and low frequency data, with the same persistence parameter.

DGP 1: Single factor linear Gaussian state space model

The data (y) and (x), and the single latent factor (F), are such that:

$$\begin{aligned} F_{t+j/3} &= \rho F_{t+(j-1)/3} + \eta_{t+j/3}, \\ y_{t+j/3}^* &= \gamma_1 F_{t+j/3} + u_{y,t+j/3}, \\ u_{y,t+j/3} &= d \cdot u_{y,t+(j-1)/3} + \sigma_y \varepsilon_{y,t+j/3}, \\ x_{t+j/3} &= \gamma_2 F_{t+j/3} + u_{x,t+j/3}, \\ u_{x,t+j/3} &= d \cdot u_{x,t+(j-1)/3} + \sigma_x \varepsilon_{x,t+j/3}, \quad t = 1, \dots, T, \quad j = 0, 1, 2, \end{aligned}$$

where the low frequency variable y is stock-sampled, and (η) , (ε_y) and (ε_x) are mutually independent i.i.d. standard Gaussian processes. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $d = 0$, $\sigma_y = \sigma_x = 1$. We consider two values for the persistence of the latent factor, that are $\rho = 0.5$ and $\rho = 0.9$.

The number of structural parameters in DGP1 is 6. In each Monte Carlo simulation, we

draw from this DGP samples of sizes $T = 100$ (corresponding to 25 years of quarterly data), $T = 200$ and $T = 500$. We perform 1000 Monte Carlo repetitions. On each simulated sample we compute the Indirect Inference (II) estimator $\hat{\theta}_{TmS}^{IIV}$ of Gouriéroux, Monfort, and Renault (1993) and Smith (1993) as described in Section 3.4, and the associated reprojections $\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{TmS}^{IIV})$ as described in Section 4. The auxiliary model is a U-MIDAS regression for the low frequency data with $\tilde{K}_x = \tilde{K}_y = 3$ and an AR(9) process for the high frequency data (see equation (3.9)). This auxiliary model has 30 parameters and yields an overidentified II setting. Instead of running S simulations from the DGP of length T , we simulate a unique long path, i.e. we set $S = 1$ and $T^S = 50000$. Moreover, we use the identity weighting matrix. The reprojection of the latent factor is computed by regression on a simulated sample of size $T^{reproj} = 100000$.

In this linear Gaussian state space model, the MLE estimator of the model parameters $\hat{\theta}_{Tm}^{ML}$ and the Kalman filter of the latent factor values - which we denote $\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{Tm}^{ML})$ - serve as the natural benchmark. We compute the Kalman filter and the ML estimates using the algorithm presented in Appendix A.

In the second Monte-Carlo design, the DGP is a two-factor linear state space model ($n_f = 2$). The two latent factors follow independent AR(1) processes, with same autoregressive parameter.

DGP 2: Two-factor linear Gaussian state space model

The data (y) and (x), and the bivariate latent factor (F), are such that:

$$\begin{aligned} F_{t+j/m} &= \begin{bmatrix} \rho & 0 \\ 0 & \rho \end{bmatrix} F_{t+(j-1)/3} + \eta_{t+j/3}, \\ y_{t+j/3}^* &= \gamma_1' F_{t+j/3} + u_{y,t+j/3}, \\ u_{y,t+j/3} &= d \cdot u_{y,t+(j-1)/3} + \sigma_y \varepsilon_{y,t+j/3}, \\ x_{t+j/3} &= \gamma_2' F_{t+j/3} + u_{x,t+j/3}, \\ u_{x,t+j/3} &= d \cdot u_{x,t+(j-1)/3} + \sigma_x \varepsilon_{x,t+j/3}, \quad t = 1, \dots, T, \quad j = 0, 1, 2, \end{aligned}$$

where the low frequency variable y is stock-sampled, and (η) , (ε_y) and (ε_x) are mutually independent i.i.d. Gaussian processes, with distribution $\mathcal{N}(0, I_2)$ for (η) , and distribution $\mathcal{N}(0, 1)$ for (ε_y) and (ε_x) . The true values of the parameters are $\rho = 0.9$, $\gamma_1 = (1, 0.2)'$,

$\gamma_2 = (0.2, 1)'$, $d = 0$, $\sigma_y = \sigma_x = 1$.

The number of structural parameters in DGP2 is 8. The sample sizes are $T = 100$, $T = 200$ and $T = 500$. We compute the II estimator $\hat{\theta}_{TmS}^{IV}$ of Gouriéroux, Monfort, and Renault (1993) and Smith (1993) and the associated reprojections $\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{TmS}^{IV})$ with the same auxiliary model and the same simulation length as for DGP1. We also compute the MLE $\hat{\theta}_{Tm}^{ML}$ and the Kalman filter estimates $\hat{F}_{t+j/m|t+j/m}(\hat{\theta}_{Tm}^{ML})$ with the algorithm in Appendix A.

The third DGP is a mixed frequency state space model with stochastic volatility. This DGP features a single Gaussian AR(1) factor in the mean of high frequency and low frequency observables ($n_f = 1$). The measurement error of the low frequency variable is a Gaussian AR(1) process. The measurement error of the high frequency variable is a conditionally heteroskedastic process.

The number of structural parameters in DGP3 is 8. The SV specification for the high frequency innovations in equations (5.1) and (5.2) is a reparametrization of the one proposed in equations (2.7) and (2.8). This specification is analogous to the one used by Monfardini (1998), Marcellino, Porqueddu, and Venditti (2016) and Clark (2011), among others. In particular, here h is the log volatility process, and is normalized to have mean zero. In this parameterization, both latent factors have a linear autoregressive dynamics.

DGP 3: Stochastic volatility model

The data (y) and (x), and the scalar latent factors (F) and (h), are such that:

$$\begin{aligned} F_{t+j/3} &= \rho F_{t+(j-1)/3} + \eta_{t+j/3}, \\ y_{t+j/3}^* &= \gamma_1 F_{t+j/3} + u_{y,t+j/3}, \\ u_{y,t+j/3} &= d \cdot u_{y,t+(j-1)/3} + \sigma_y \varepsilon_{y,t+j/3}, \\ x_{t+j/3} &= \gamma_2 F_{t+j/3} + \sigma_x \exp \left\{ \frac{1}{2} h_{t+j/3} \right\} \varepsilon_{x,t+j/3}, \end{aligned} \tag{5.1}$$

$$h_{t+j/3} = \rho_{SV} h_{t+(j-1)/3} + \nu \cdot \xi_{t+j/3}, \quad t = 1, \dots, T, \quad j = 0, 1, 2, \tag{5.2}$$

where the low frequency variable y is stock-sampled, and (η) , (ε_y) , (ε_x) and (ξ) are mutually independent i.i.d. standard Gaussian processes. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $d = 0$, $\sigma_y = \sigma_x = 1$, $\rho_{SV} = 0.95$, $\nu = 0.3$. We consider two values for the persistence of the latent factor in the conditional mean, that are $\rho = 0.5$ and $\rho = 0.9$.

Again, the sizes of the simulated samples are $T = 100$, $T = 200$ and $T = 500$. We have tried different auxiliary models for the indirect inference procedure, including GARCH(1,1) for the squared high frequency residuals, an AR(10) model on the logarithm of the squared high frequency residuals and an AR(10) model on the logarithm of the squared high frequency observables (similarly as in Monfardini (1998)). Barigozzi, Halbleib-Chiriac, and Veredas (2014) show that the GARCH(1,1) model is the best auxiliary model for estimating a stochastic volatility model with Indirect Inference, in the sense that it provides the best trade-off between efficiency and estimation noise. The GARCH(1,1) auxiliary model reduces, however, the computational speed of the indirect inference estimator, as it requires estimation via maximum likelihood. We therefore prefer an AR-ARCH specification in the auxiliary model, since this allows for estimation via a simple two-step approach based on OLS regressions. Specifically, we compute the indirect inference estimator $\hat{\theta}_{TmS}^{IV}$ of Gouriéroux, Monfort, and Renault (1993) and Smith (1993) as described in Section 3.4, using the auxiliary model in equation (3.10), with $\tilde{K}_x = \tilde{K}_y = 4$ in the U-MIDAS regression for the low frequency data, and an AR(9)-ARCH(10) specification for the high frequency data.⁸ We compare the distribution of our indirect inference estimator with the distribution of the MLE in the Monte Carlo simulations. In the nonlinear state space model of DGP 3, we implement the MLE via a simulation-based EM algorithm as described in Section 3.5. See Appendix B for the detailed algorithm.

Note that for our DGPs we have made the modeling choices of (i) not having multiple observable high frequency and low frequency variables, and (ii) not including a new equation for the volatility dynamics of the DGP for the latent factor F_t . These simplifying hypotheses allow to keep at a minimum the number of parameters to be estimated, easing the implementation and the comparison of the different estimation methodologies, as they avoid numerical convergence issues.⁹

⁸ In this paper we do not consider the moment matching procedure of Gallant and Tauchen (1996). However, adopting their procedure, which is computationally even more attractive, could make the use of GARCH-type auxiliary models more attractive. As the Gallant-Tauchen procedure is based on the score, it would not require iterated ML estimates (see for instance, Sentana, Calzolari, and Fiorentini (2008)).

⁹The setup of the model is simplified so that the II criteria in equations (3.12) and (3.13) are functions of a relatively small number of true parameters (6 to 8, depending of the DGP). Therefore a global minimum for these objective functions can be easily found numerically using standard (modifications of) Quasi-Newton algorithms, as the one provided by Matlab’s “fminunc” function, adapted to take into account the domain constraints of each parameter. The fact that the optimization of the objective functions does not create numerical issues, and is fast to perform, allows us to better understand the finite-sample properties of the II estimators in the MC application, and minimizes the computing time of an estimate. The problems of including multiple HF (or LF) observables in the same model, or adding stochastic volatility in the unobserved

5.2 Monte Carlo results in the linear Gaussian state space model

Tables 1 through 3 report the results for the linear Gaussian state space models in DGP1 and DGP2. For each combination of DGP parameters and sample size, we provide the results of the Indirect Inference (II) procedure for parameter estimation and filtering of the latent factor path. As a benchmark, we also provide the estimation and filtering results using the Maximum Likelihood (ML) procedure based on the Kalman filter.¹⁰

In Table 1, we consider DGP1 where the single latent factor process is mildly persistent with autocorrelation $\rho = 0.5$. The finite sample performance of the II estimator is remarkably good. First, it has only a small bias for most configurations. Second - as expected - the ML estimator based on the Kalman filter is more efficient, but the efficiency loss of the II estimator is rather limited. The bias of both the II and MLE is more pronounced for parameter σ_y , that is the volatility of the low frequency measurement error. For this parameter, the efficiency loss of the II estimator compared to MLE is a bit larger. As expected, the dispersions of the estimators decrease with the sample size T . Moreover, the reprojection procedure provides rather accurate estimates of the latent factor values. Indeed, the average correlation between true and filtered factor values is about 0.80 for all sample sizes, which is close to the performance of the Kalman filter.

In unreported MC results we compared the performance of the above II estimator - which uses the U-MIDAS/AR auxiliary model for high/low frequency data - with the performance of the II estimator using only the low-frequency U-MIDAS specification as auxiliary model. The II estimator of the standard deviation parameter for the high-frequency data σ_x based on the low-frequency U-MIDAS auxiliary model has a large bias. As shown in Table 1, this problem does not arise when we include high-frequency data in the auxiliary model via the mixed frequency VAR specification. These findings confirm the intuition that using data at both frequencies provides a more informative auxiliary model.

In Table 2, the autocorrelation of the latent factor in DGP1 is set equal to $\rho = 0.9$. Both the ML and II estimators have smaller dispersions in this MC design compared to Table

factor dynamics, is only of computational nature, and not a theoretical limitation. In this case a numerically more efficient optimization routine for the same II objective function could be used, as the ones proposed in Chernozhukov and Hong (2003), and more recently by Creel, Gao, Hong, and Kristensen (2015).

¹⁰All Monte Carlo simulations in Section 5 have been performed using Matlab 7.10.0 (R2010a) on a laptop with a 1.60 GHz processor and 4 GB RAM. Optimization problems involved in parameter estimation have been solved using the Matlab procedure ‘fminunc’.

1. This effect is due to the more favorable signal-to-noise setting when ρ is changed from 0.5 to 0.9 in our parameterization of the DGP. Indeed, with $\rho = 0.9$ the factor has a larger unconditional variance relative to the noise variance, which is fixed across the two cases. Hence, the signal-to-noise ratio is larger for the DGP in Table 2 compared to Table 1.

In Table 3 we report the simulation results for DGP2, which features two latent factors, with loadings equal to $\gamma_1 = (1.0, 0.2)'$ and $\gamma_2 = (0.2, 1.0)'$. Compared to the one-factor case in Tables 1 and 2, the loadings are estimated rather precisely, with the dispersion of the loadings equal to 0.2 being larger than that of the loadings equal to 1.0. Also in this case we find that the II estimator has a very good performance, with the exception of the estimator of the low frequency volatility σ_y , which has a bias of around 20% for small sample sizes ($T = 200, 100$), and a large dispersion. Nevertheless, the reprojection procedure produces accurate estimates of both factors. As expected, the factor which loads mainly on the high frequency observables (that is F_2) is estimated more precisely (average correlation with the true factor equal to 0.88 for $T = 100$) than the factor which loads mainly on the low frequency observables (average correlation with the true factor equal to 0.74 for $T = 100$).

Overall, the results in Tables 1 through 3 are remarkably impressive, since they show that the performance of the II procedure is rather close to the efficient benchmark in the linear Gaussian state space model.

5.3 Monte Carlo results for the state space stochastic volatility model

We now consider the more challenging state space model with stochastic volatility in DGP3. Tables 4 and 5 report the results of Monte Carlo simulations comparing the II estimator with the MLE (implemented via a simulation-based EM algorithm) for sample sizes $T = 500$ and $T = 200$ respectively. These tables compare the two estimation methods on the basis of (a) statistical criteria - mean/bias/quantiles of sampling distributions, (b) filtering accuracy - both for conditional mean and volatility factors, and (c) computational time. Compared to the linear Gaussian state space model in DGP1, the structural model now has two additional parameters, which are the autoregressive coefficient ρ_{SV} and the volatility parameter ν of the log stochastic volatility process. A first encouraging finding is that the estimation results for parameters $\gamma_1, \gamma_2, \rho, d, \sigma_y, \sigma_x$ are comparable to those of the Gaussian state space model displayed in Tables 1 and 2, with slightly larger dispersions in Tables 4 and 5, as expected.

The latter effect is more pronounced for parameter d , the autoregressive coefficient of the low frequency measurement error, for both the II estimator and the MLE. The stochastic volatility parameters ρ_{SV} and ν are estimated with rather small biases. Note that sample size $T = 200$ corresponds to 600 high frequency observations, and for such sample sizes the estimation of ARCH and SV specifications can be inaccurate, even in absence of latent factors in the mean. Yet, comparing the distributions of II and ML estimates, we observe that also in the stochastic volatility case the efficiency loss of the former estimator is limited.

It is worth noting that the reprojection method provides rather accurate estimates of the latent factor values also in the stochastic volatility model. Results are less good for the log volatility factor (average correlation between estimated and true factor values equal to 0.55 for sample size $T = 500$ in the design with $\rho = 0.5$). This result is not surprising, because there is no obvious choice for the transformations of the observable variables, whose linear combination provides the best approximation of the conditional expectation of the volatility factor in this nonlinear state space model. In Tables 4 and 5 we use current and past values of log squared high frequency residuals, but other choices could yield better results.

The II procedure provides a substantial reduction in computational time compared to the simulation-based EM procedure used to obtain the ML estimates. For instance, the computation of the II estimates for one Monte Carlo repetition in the stochastic volatility design with $\rho = 0.5$ and sample size $T = 200$ takes on average about 18 minutes, against the 24 minutes required on average for the ML estimates. The difference is larger with sample size $T = 500$, for which the average computational times are 16 minutes for II and 61 minutes for ML. Here, the computational time for the II procedure is less than 21 minutes in 75% of the MC replications, while the computational time for ML is more than one hour in more than 25% of the MC replications. Sample sizes such as $T = 500$ or even larger are often encountered in financial datasets, if the lower frequency is weekly or monthly.

To summarize the findings of the MC simulations with the stochastic volatility design, the II procedure offers a substantial gain in computational time compared to the ML procedure implemented via Monte Carlo EM, while the cost in terms of efficiency loss is limited.

6 Empirical study

We present an empirical application of our model to the problem of forecasting at short horizons the Euro-area quarterly GDP growth using monthly macroeconomic indicators.

6.1 Data and model specification

The dataset is the same as the one considered in the empirical study of Marcellino, Porqueddu, and Venditti (2016).¹¹ The data consists of the quarterly GDP growth rates for the Euro-area (GDP) observed from 1991-Q1 to 2011-Q1, and the monthly observations for the same period, i.e. from 1991-M1 to 2011-M3, for the following 8 macroeconomic indicators: (1) the aggregate European Industrial Production index for all sectors of the European economy: IP, (2) the European Industrial Production index for “Pulp and Paper sector”: IP-Pulp/Paper, (3) the Germany IFO Business Climate Index: IFO, (4) the Euro-area Economic Sentiment Index: ESI, (5) the Euro-area Composite Purchasing Manager Index: PMI, (6) the bilateral dollar-euro exchange rate, measured as year-on-year percentage growth: EXC, (7) the difference between 3-month and 10-year US Treasury bond yield: SPR, and (8) the University of Michigan consumer sentiment index for the US: MICH. In line with the empirical study of Bai, Ghysels, and Wright (2013), we consider the first difference of the series (3) to (8) to induce stationarity, and we normalize all series by their full sample mean and standard deviation.¹²

We estimate the mixed-frequency stochastic volatility model defined as DGP 3 in Section 5 and the linear Gaussian factor model defined as DGP 1 on eight different pairs of mixed frequency observables.¹³ In each model we include GDP as the low frequency observable, and one of the eight monthly indicators listed above as the high frequency variable. We assume the presence of one high frequency latent factor ($n_f = 1$), and that the observed quarterly GDP is the sum of three unobservable monthly growth rates: $y_t = y_t^* + y_{t-1/3}^* + y_{t-2/3}^*$. Thus, we have $m = 3$ and the low frequency variable is flow sampled. We estimate the SV model by the Indirect Inference (II) procedure, using the same auxiliary model as in the MC simulations of Section 5, and deploy the II estimates in the reprojection procedure to filter

¹¹We thank M. Marcellino, M. Poqueddu and F. Venditti for sharing their dataset with us.

¹²Augmented Dickey-Fuller tests failed to reject the null hypothesis of a unit root for series (3) to (8).

¹³Differently from the specification of DGP 1 in Section 5, we allow the autoregressive parameters d_x and d_y of the idiosyncratic error terms u_x and u_y to be different.

the latent factors. We estimate the Gaussian state space model without SV by adapting the Kalman filter for periodic state space models proposed in Bai, Ghysels, and Wright (2013) to accommodate flow sampling (see Appendix C).

6.2 Estimation and in-sample explanatory power

Before performing the forecasting exercise, we discuss the estimation results of the models for the entire data sample ending in 2011-Q1. In Table 6 we report the values of the R^2 of the regression of both GDP and the five monthly indicators (1)-(5) on the filtered values of the latent factor F in each model.¹⁴ For all five considered models, the factor explains a substantial fraction of the variability of both the GDP and the respective HF monthly indicator. For the factor model with stochastic volatility estimated with the IP indicator, the common factor explains 74% of the variability of GDP and 48% of Industrial Production. When the factor model without SV is estimated on the same data, the explanatory power of the common factor for GDP is slightly higher, as the R^2 increases to 82%. On the other hand, the explanatory power for GDP (resp. the HF indicator) of the factor extracted using the IFO and ESI survey indices, are higher (resp. lower) for the SV model than for the linear Gaussian one.¹⁵ The factor extracted using the IP-Pulp/Paper index explains only 11% of the variability of this HF index for both models, but this is not surprising as the Pulp and Paper sector represents a small fraction of the total Industrial Production. Moving to the estimation of our model using the EXC, SPR and MICH indicators, both in the full sample and in the shorter subsamples considered in the forecasting exercise below, the regressions produce loadings of the HF observables on the factor close to zero, and a filtered factor uncorrelated with the corresponding HF variable, with no forecasting power for GDP. For this reason we report only results for the five monthly indicators (1)-(5). It should be noted

¹⁴The regression of GDP on the factor is a special type of MIDAS regression, in which we regress the value of GDP growth at the end of quarter t on the sum of the filtered values of the factor in the months of the same quarter:

$$\begin{aligned} y_t &= y_t^* + y_{t-1/3}^* + y_{t-2/3}^* \\ &= \gamma_1(F_t + F_{t-1/3} + F_{t-2/3}) + u_{y,t} + u_{y,t-1/3} + u_{y,t-2/3}. \end{aligned}$$

On the other hand, each high frequency indicator is regressed only on the contemporaneous value of the factor. In Table 6 we report R^2 instead of the values of the loadings of the factor on observables, as they are more easily interpretable.

¹⁵These results are robust to the choice of the starting point of the estimation algorithms, which did not show convergence problems for the series reported in the tables.

that our mixed frequency model admits only the contemporaneous impact of the common latent factor on the HF variable, and the impact of the factor values within a quarter on the flow-sampled LF variable. It could be that more general specifications, such as a factor model in which the observables load on more lags of the latent factor - on the last 12 months, for instance - might be more appropriate to assess the forecasting power for the European GDP of the 2 US macroeconomic indicators SPR and MICH, and the Euro-dollar exchange rate EXC.¹⁶

Figures 1 to 5 display the time series of the observable variables used to estimate the factor models, and the filtered mean and stochastic volatility factor paths obtained via reprojection, corresponding to HF indicators (1) to (5). Visual inspection of the estimated factor paths \hat{F} of the five figures reveal commonalities across models, like the major drop and the successive rebound following the financial crisis of 2008. Nevertheless, the relative size of this drop appears to be more pronounced for the two IP series than for IFO, ESI and PMI.¹⁷ The trajectories of the filtered stochastic volatility factor are represented in Panel (d) of Figure 1, and Panels (c) of Figures 2 to 5. For all but one of the monthly indicators, the estimated idiosyncratic volatility factor oscillates around zero before 2008 and then increases to values larger than 1.5, indicating that the idiosyncratic volatility of the monthly macroeconomic series more than doubled during the recent financial crisis.¹⁸ Only the IP-Pulp/Paper idiosyncratic volatility shows a different behavior, being much larger in the first half of the sample, than in the second one. We stress that we do not impose any dependence structure between the mean factor F and the stochastic volatility factor h specific to each HF series, and this fact might be relevant for the situations like the one of the IP-Pulp/Paper monthly indicator in which the large drop of the mean factor in 2009, corresponding to the drop in DGP, is not associated with a spike in the volatility of the high frequency index.

¹⁶See Marcellino, Porqueddu, and Venditti (2016), in particular Section B of their online Appendix, for an example of a richer dependence structure between the observables and the factor. Nevertheless, this result is not surprising as the loadings of EXC, SPR and MICH on their common latent factor summarizing the current state of the business cycle, are much smaller - in absolute value - than the loadings of the other five macroeconomic indicators.

¹⁷As the estimated loadings of the latent factor F on the observables have positive signs, a drop in the factor is associated with a drop in both GDP and the monthly indicator in the same quarter.

¹⁸Indeed, the value of volatility $\exp(0.5 \cdot h_t)$ increases from 1 to more than 2.1, when factor h_t goes from 0 to 1.5.

6.3 Forecasting

As the in-sample estimates of our five factor models are different, we expect the models to have different forecasting power for the GDP. Similarly to Marcellino, Porqueddu, and Venditti (2016), we perform an out-of-sample forecasting exercise where at the end of each quarter we estimate the models with and without stochastic volatility, and use them to forecast GDP up to an horizon of $H = 4$ quarters ahead of the estimation sample final date. The first estimation window is from 1991-Q1 to 2005-Q4, and is recursively expanded up to 2010-Q4. In Table 7 we report the Root Mean Squared Forecasting Errors (RMSFE) as ratios to the RMSFE of a forecasting model assuming constant growth of the GDP. An entry below one in Table 7 indicates that the factor model outperforms the naive constant growth benchmark. This choice allows us to have comparable results across different models, forecasting horizons, but also with the results of Marcellino, Porqueddu, and Venditti (2016, Figures 8 and 9).

We immediately note that the forecasting ability of all models, relative to the naive benchmark, is limited to short horizons up to 2 quarters ahead. Indeed, all the RMSFE ratios reported in Table 7 are very close to, or even larger than, 1 for forecasting horizons $H = 3, 4$ quarters. Note that Marcellino, Porqueddu, and Venditti (2016) report the RMSFE ratios for a maximum of 7 months ahead, and that the RMSFE for 6 months (i.e. $H = 2$ quarters) is always very close to 1 for all their models. For the factor models estimated using the aggregate Industrial Production index, in Table 7 the linear Gaussian model seems to outperform the model incorporating SV when used to forecast GDP at 1 quarter horizon, as the RMSFE ratio for the latter model is 0.7, which is smaller than the value slightly below 0.8 reported by both our SV model, and by Marcellino, Porqueddu, and Venditti (2016) for all their specifications. On the other hand, the results are completely different when considering the 1 quarter ahead forecasting accuracy of our SV models estimated on the IFO and ESI indexes (RMSFE around 0.7 for $H = 1$, and 0.9 for $H = 2$, in both models), which clearly outperform our models without SV (only the IFO model has a RMSFE lower than 1, equal to 0.9 for $H = 1$) and the model of Marcellino, Porqueddu, and Venditti (2016) (RMSFE around 0.8 for $H = 1$, and around 1.0 for $H = 1$) at both 1 and 2 quarters ahead horizons. Finally, the models estimated on IP-Pulp/paper and PMI show some forecasting power at 1 quarter horizon, yet with larger RMSFE compared to all models discussed above.

Overall, the results of this empirical exercise demonstrate the importance of considering

stochastic volatility when estimating mixed frequency factor models both for the in-sample explanatory power of the extracted factors, which might be important when constructing coincident indexes of the economy as in Marcellino, Porqueddu, and Venditti (2016), and for the out-of sample predictive ability of the estimated model. Moreover, the estimation of our SV models on GDP (LF series) and only one monthly macroeconomic indicator (HF series) showed, that the forecasting accuracy of the different macroeconomic variables can be different across different variables, horizons and model specifications.

There is scope for even further improvements - despite the fact that some of our models already outperform the approach suggested by Marcellino, Porqueddu, and Venditti (2016) - using the same data and sample configurations. In our approach, we followed Bai, Ghysels, and Wright (2013) who focused exclusively on bivariate specifications, whereas Marcellino, Porqueddu, and Venditti (2016) build one joint model for the eight series considered. We have in principle 8 forecasts obtained from the paired bivariate models - with some outperforming and some mostly at par with the single large model they consider. In light with Andreou, Ghysels, and Kourtellis (2013) we could further improve the forecasting output by constructing forecast combinations of our 8 predictions - ultimately producing a single combination forecast. Since the scope of our paper is not to produce the best forecasting model, but rather show the possibilities of estimating and implementing state space models with SV using a new indirect inference approach, we refrain from adding these further improvements.

Finally, the procedures we implemented lend themselves easily to nowcasting simply by adopting a MIDAS with leads regression approach - see Andreou, Ghysels, and Kourtellis (2013) for further details. As noted in Section 4, this is only done at the reprojection stage. Hence, the model parameter estimates suffice to run another simulation to obtain the nowcasting models.

7 Conclusions

We proposed a fairly simple and remarkably accurate indirect inference estimation procedure for state space models with either Gaussian errors or stochastic volatility. We consider a mixed frequency data setting as it is a typical situation where stochastic volatility is relevant due to the use of high frequency data. We confined our attention to settings involving only a

single high and low frequency data series. Yet, the methods can easily be extended to more series of either type as the mixed frequency VAR auxiliary model can straightforwardly accommodate such settings. A more challenging extension involves larger values of m - the differences in low and high frequencies. The use of U-MIDAS regressions makes our approach extremely computationally attractive due to the use of OLS. With larger values of m we know that U-MIDAS becomes over-parameterized. While regular MIDAS regressions are a feasible alternative - they require non-linear estimation and are therefore less appealing. It should also be noted that we only covered indirect estimation procedures. It would also be fairly straightforward to apply the moment matching procedure of Gallant and Tauchen (1996) instead. As is well known, this would make our procedures potentially computationally even more attractive, while maintaining the same asymptotic properties. This would also broaden the potential set of auxiliary models, including GARCH and EGARCH, as the Gallant and Tauchen (1996) procedure is based on the empirical score and does not require repeated ML estimates. An interesting extension in this regard would be to use the criteria introduced by Barigozzi, Halbleib-Chiriac, and Veredas (2014) for choosing the best auxiliary model.

Last but not least, it should be noted that we assumed that the number of factors is known. In practice, one should of course also consider testing for the number of factors. There is a considerable literature on testing for the number of factors. In terms of testing, it is worth noting that the indirect inference procedures should not pose any additional issues in terms of testing the number of factors. See in particular Guay and Scaillet (2003) who study a hypothesis testing problem quite similar to determining the number of factors - namely involving unidentified parameters under the null - in the context of indirect inference.

References

- Andreou, Elena, Eric Ghysels, and Andros Kourtellis, 2010, Regression models with mixed sampling frequencies, *Journal of Econometrics* 158, 246–261.
- , 2011, Forecasting with mixed-frequency data, in Michael P Clements, and David F Hendry, ed.: *Oxford Handbook of Economic Forecasting* pp. 225–245.
- , 2013, Should macroeconomic forecasters use daily financial data and how?, *Journal of Business and Economic Statistics* 31, 240–251.
- Aruoba, S Borağan, Francis X Diebold, and Chiara Scotti, 2009, Real-time measurement of business conditions, *Journal of Business and Economic Statistics* 27, 417–427.
- Assimakis, N., and M. Adam, 2009, Steady state Kalman filter for periodic models: A new approach, *International Journal of Contemporary Mathematical Sciences* 4, 201–218.
- Bai, Jennie, Eric Ghysels, and Jonathan H. Wright, 2013, State space models and MIDAS regressions, *Econometric Reviews* 32, 779–813.
- Banbura, Marta, Domenico Giannone, Michele Modugno, and Lucrezia Reichlin, 2013, Nowcasting and the real-time data flow, ECB working paper.
- Banbura, Marta, and Michele Modugno, 2014, Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data, *Journal of Applied Econometrics* 29, 133–160.
- Barigozzi, Matteo, Roxana Halbleib-Chiriac, and David Veredas, 2014, Which model to match?, Available at SSRN 1986419.
- Bernanke, Ben, Mark Gertler, and Mark Watson, 1997, Systematic monetary policy and the effects of oil price shocks, *Brookings Papers on Economic Activity* 1, 91–157.
- Carriero, Andrea, Todd Clark, and Massimiliano Marcellino, 2012, Common drifting volatility in large Bayesian VARs, CEPR Discussion Paper No. 8894.
- , 2013, Real-time nowcasting with large bridging equations with time-varying volatility, CEPR Discussion Paper No. 9312.

- Carvalho, Carlos M., Michael S. Johannes, Hedibert F. Lopes, and Nicholas G. Polson, 2010, Particle learning and smoothing, *Statistical Science* 25, 88–106.
- Chan, K. S., and J. Ledolter, 1995, Monte Carlo EM Estimation for Time Series Models Involving Counts, *Journal of the American Statistical Association* 90, 242–252.
- Chen, Xilong, and Eric Ghysels, 2010, News - good or bad - and its impact on volatility predictions over multiple horizons, *Review of Financial Studies* pp. 46–81.
- Chernozhukov, V., and H. Hong, 2003, MCMC approach to classical estimation, *Journal of Econometrics* 115.
- Clark, Todd E, 2011, Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility, *Journal of Business and Economic Statistics* 29, 327–341.
- Cox, D.R., 1981, Statistical analysis of time series: Some recent developments [with discussion and reply], *Scandinavian Journal of Statistics* 8, 93–115.
- Creal, Drew, 2012, A survey of sequential Monte Carlo methods for economics and finance, *Econometric Reviews* 31, 245–296.
- Creel, M., J. Gao, H. Hong, and D. Kristensen, 2015, Bayesian Indirect Inference and the ABC of GMM, *Working Paper*.
- Dempster, A.P., N.M. Laird, and D.B. Rubin, 1977, Maximum Likelihood from incomplete data via di EM algorithm, *Journal of the Royal Statistical Society Series B* 39, 1–38.
- Douc, R., E. Moulines, and J. Olsson, 2009, Optimality of the auxiliary particle filter, *Probability and Mathematical Statistics* 29, 1–28.
- Doucet, A., Johansen A., 2010, A tutorial on particle filtering and smoothing: fifteen years later, in D. Crisan, and B. Rozovsky, ed.: *Handbook of Nonlinear Filtering* (Oxford University Press: Oxford).
- Engle, R. F., and Gary G Lee, 1999, A permanent and transitory component model of stock return volatility, in R.F. Engle, and H. White, ed.: *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger* pp. 475–497. Oxford University Press, Oxford.

- Foroni, Claudia, and Massimiliano Marcellino, 2013, A survey of econometric methods for mixed-frequency data, Available at SSRN 2268912.
- , and Christian Schumacher, 2013, Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials, *Journal of the Royal Statistical Society: Series A*.
- Gallant, A. Ronald, and George Tauchen, 1996, Which moments to match?, *Econometric Theory* 12, 657–681.
- , 1998, Reprojecting partially observed systems with application to interest rate diffusions, *Journal of the American Statistical Association* 93, 10–24.
- Ghysels, Eric, 2016, Macroeconomics and the reality of mixed frequency data, *Journal of Econometrics* 193, 294–314.
- Ghysels, E., A. Harvey, and E. Renault, 1996, Stochastic volatility, in G.S. Maddala, and C.R. Rao, ed.: *Handbook of Statistics, Volume 14 - Statistical Methods in Finance*. North Holland.
- Ghysels, Eric, Jonathan B Hill, and Kaiji Motegi, 2014, Regression-based mixed frequency Granger causality tests, Discussion paper, Technical report, Department of Economics, University of North Carolina, Chapel Hill.
- , 2016, Granger causality in mixed frequency vector autoregressive models, *Journal of Econometrics* - forthcoming.
- Ghysels, Eric, Pedro Santa-Clara, and Rossen Valkanov, 2006, Predicting volatility: getting the most out of return data sampled at different frequencies, *Journal of Econometrics* 131, 59–95.
- Ghysels, Eric, and Jonathan Wright, 2009, Forecasting professional forecasters, *Journal of Business and Economic Statistics* 27, 504–516.
- Giannone, Domenico, Lucrezia Reichlin, and David Small, 2008, Nowcasting: The real-time informational content of macroeconomic data, *Journal of Monetary Economics* 55, 665–676.
- Godsill, S., A. Doucet, and M. West, 2004, Monte Carlo smoothing for non-linear time series, *Journal of the American Statistical Association* 199, 156–168.

- Götz, Thomas B, and Alain Hecq, 2014a, Nowcasting causality in mixed frequency vector autoregressive models, *Economics Letters* 122, 74–78.
- , 2014b, Testing for Granger Causality in Large Mixed-Frequency VARs, Discussion paper, Technical report, Department of Economics, University of Maastricht.
- Gouriéroux, C., and A. Monfort, 1997, *Simulation-based econometric methods* (Oxford University Press).
- Gouriéroux, C., A. Monfort, and E. Renault, 1993, Indirect inference, *Journal of Applied Econometrics* 8, 85–118.
- Guay, Alain, and Olivier Scaillet, 2003, Indirect inference, nuisance parameter, and threshold moving average models, *Journal of Business and Economic Statistics* 21, 122–132.
- Hamilton, James D., 1994, State-space models, in R.F. Engle, and D.L. McFadden, ed.: *Handbook of Econometrics - Volume 4* pp. 3039–3080. Elsevier.
- Harvey, Andrew, 1989, *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, Cambridge).
- Harvey, Andrew C., and Richard G. Pierse, 1984, Estimating missing observations in economic time series, *Journal of the American Statistical Association* 79, 125–131.
- Kantas, N., A. Doucet, S. S. Singh, J. M. Maciejowski, and N. Chopin, 2015, On particle methods for parameter estimation in general state-space models, *Statistical Science* 30, 328–351.
- Kim, J., and D. Stoffer, 2008, Fitting stochastic volatility models in the presence of irregular sampling via particle methods and the EM algorithm, *Journal of Time Series Analysis* 29, 811–833.
- Kitagawa, G., 1996, Monte carlo filter and smoother for non-gaussian nonlinear state-space models, *Journal of Computational and Graphical Statistics* 5, 1–25.
- , and S. Sato, 2001, Sequential Monte Carlo Methods in Practice, in *Sequential Monte Carlo Methods in Practice* (New York: Springer-Verlag).

- Kuzin, Vladimir, Massimiliano Marcellino, and Christian Schumacher, 2011, MIDAS versus mixed-frequency VAR: Nowcasting GDP in the Euro area, *International Journal of Forecasting* 27, 529–542.
- Lopes, H. F., and R. Tsay, 2011, Particle filters and Bayesian inference in financial econometrics, *Journal of Forecasting* 30, 168–209.
- Marcellino, M., M. Porqueddu, and F. Venditti, 2016, Short-term GDP forecasting with a mixed-frequency dynamic factor model with stochastic volatility, *Journal of Business and Economic Statistics* 34, 118–127.
- Marcellino, Massimiliano, and Christian Schumacher, 2010, Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP, *Oxford Bulletin of Economics and Statistics* 72, 518–550.
- Mariano, R.S., and Y. Murasawa, 2003, A new coincident index of business cycles based on monthly and quarterly series, *Journal of Applied Econometrics* 18, 427–443.
- Mitnik, S., and P. Zadrozny, 2005, Forecasting quarterly German GDP at monthly intervals using monthly Ifo business conditions data, in J.-E. Sturm, and Wollmershäuser. T., ed.: *Ifo Survey Data in Business Cycle and Monetary Policy Analysis*. Springer.
- Monfardini, Chiara, 1998, Estimating stochastic volatility models through indirect inference, *The Econometrics Journal* 1, 113–128.
- Neath, Ronald C., 2013, On convergence properties of the Monte Carlo EM algorithm, *Institute of Mathematical Statistics Collections* 10, 43–62.
- Nunes, L.C., 2005, Nowcasting quarterly GDP growth in a monthly coincident indicator model, *Journal of Forecasting* 24, 575–592.
- Olsson, J., O. Cappé, R. Douc, and E. Moulines, 2008, Sequential Monte Carlo smoothing with application to parameter estimation in nonlinear state space models, *Bernoulli* 14, 155–179.
- Pastorello, Sergio, Eric Renault, and Nizar Touzi, 2000, Statistical inference for random-variance option pricing, *Journal of Business and Economic Statistics* 18, 358–367.
- Pitt, Michael K., and Neil Shephard, 1999, Filtering via simulation: Auxiliary particle filters, *Journal of the American Statistical Association* 94, 590–599.

- Sandmann, Gleb, and Siem J. Koopman, 1998, Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics* 87, 271–301.
- Schorfheide, Frank, and Dongho Song, 2013, Real-time forecasting with a mixed-frequency VAR, Discussion paper, National Bureau of Economic Research.
- Sentana, Enrique, Giorgio Calzolari, and Gabriele Fiorentini, 2008, Indirect estimation of large conditionally heteroskedastic factor models, with an application to the dow 30 stocks, *Journal of Econometrics* 146, 10–25.
- Shephard, Neil, 2005, *Stochastic Volatility: Selected Readings* (Oxford University Press).
- Smith, Anthony A, 1993, Estimating nonlinear time-series models using simulated vector autoregressions, *Journal of Applied Econometrics* 8, 63–84.
- Zadrozny, Peter, 1988, Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies, *Econometric Theory* 4, 108–124.
- Zadrozny, P.A., 1990, Forecasting US GNP at monthly intervals with an estimated bivariate time series model, *Federal Reserve Bank of Atlanta Economic Review* 75, 2–15.

Table 1: MC simulations for the single-factor linear Gaussian state space model (persistence parameter of the latent factor $\rho = 0.5$)

T = 500		MLE (Kalman filter)					Indirect Inference (Auxiliary model: U-MIDAS / AR)				
Coeff.		mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.
γ_2		0.99	-0.01	0.94	0.98	1.06	0.98	-0.02	0.86	0.97	1.10
γ_1		1.01	0.01	0.94	1.00	1.06	1.00	0.00	0.90	0.97	1.05
ρ		0.50	0.00	0.48	0.50	0.52	0.50	0.00	0.47	0.50	0.54
d		-0.01	-0.01	-0.07	0.01	0.05	-0.02	-0.02	-0.10	0.01	0.07
σ_x		1.00	-0.00	0.92	1.01	1.07	0.99	-0.01	0.87	1.04	1.15
σ_y		0.95	-0.05	0.90	0.99	1.06	0.91	-0.09	0.87	1.02	1.11
$corr(\hat{F}, F)$		0.80	-	0.79	0.80	0.81	0.78	-	0.78	0.79	0.80
T = 200											
γ_2		0.99	-0.01	0.86	1.01	1.12	0.94	-0.06	0.75	0.93	1.09
γ_1		1.03	0.03	0.91	1.02	1.13	1.01	0.01	0.85	0.94	1.16
ρ		0.50	-0.00	0.45	0.50	0.54	0.51	0.01	0.45	0.51	0.58
d		-0.03	-0.03	-0.13	-0.00	0.09	0.00	0.00	-0.13	0.05	0.14
σ_x		0.97	-0.03	0.84	1.00	1.12	1.00	0.00	0.85	1.07	1.23
σ_y		0.86	-0.14	0.79	0.97	1.05	0.79	-0.21	0.50	1.00	1.12
$corr(\hat{F}, F)$		0.79	-	0.78	0.80	0.81	0.77	-	0.75	0.78	0.79
T = 100											
γ_2		0.97	-0.03	0.82	0.96	1.12	0.89	-0.11	0.69	0.85	1.11
γ_1		1.01	0.01	0.88	0.99	1.11	0.92	-0.08	0.69	0.89	1.11
ρ		0.50	-0.00	0.43	0.50	0.59	0.51	0.01	0.40	0.54	0.67
d		-0.04	-0.04	-0.20	0.01	0.11	-0.01	-0.01	-0.21	0.07	0.18
σ_x		0.95	-0.05	0.77	1.00	1.16	0.98	-0.02	0.76	1.11	1.26
σ_y		0.81	-0.19	0.67	0.94	1.08	0.84	-0.16	0.54	1.05	1.17
$corr(\hat{F}, F)$		0.78	-	0.75	0.78	0.81	0.75	-	0.73	0.76	0.78

This table reports mean, bias, and 25%, 50%, 75% quantiles of the distribution of the ML (left) and Indirect Inference (II, right) estimators in 1000 MC replications. The data generating process is DGP1 in Section 5.1, corresponding to a mixed frequency linear state space model with a single AR(1) latent factor, $m = 3$, and stock sampling of the low frequency variable. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $\rho = 0.5$, $d = 0$, $\sigma_y = \sigma_x = 1$. The simulated samples have size $T = 500$ (top), $T = 200$ (middle), $T = 100$ (bottom). The auxiliary model for the indirect inference estimator is a U-MIDAS regression for low frequency data with $\tilde{K}_x = \tilde{K}_y = 3$ and an AR(9) model for the high frequency data (see equation (3.9)), with the correlation between the errors of the two equations as a free auxiliary parameter. The Indirect Inference estimator uses a single long simulated sample of the structural model ($S = 1$ and $T^S = 50000$) and an identity weighting matrix. We also compute the mean and 25%, 50%, 75% quantiles of the sample correlation between the estimated and true factor paths. The estimated factor paths are obtained by Kalman filter with the ML estimate (left) and the reprojection method with the II estimate (right), using $T^{reproj} = 100000$.

Table 2: MC simulations for the single-factor linear Gaussian state space model (persistence parameter of the latent factor $\rho = 0.9$)

T = 500		MLE (Kalman filter)				Indirect Inference (Auxiliary model: U-MIDAS / AR)					
		Coeff.	mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median
γ_2		1.00	-0.00	0.98	1.00	1.02	0.98	-0.02	0.94	0.97	1.02
γ_1		1.00	-0.00	0.96	1.00	1.03	0.98	-0.02	0.91	0.98	1.04
ρ		0.90	0.00	0.89	0.90	0.91	0.90	0.00	0.87	0.91	0.93
d		-0.00	-0.00	-0.03	-0.00	0.03	0.02	0.02	-0.02	0.02	0.07
σ_x		1.00	-0.00	0.96	1.00	1.03	1.02	0.02	0.98	1.04	1.08
σ_y		0.99	-0.01	0.96	1.00	1.03	1.03	0.03	0.97	1.04	1.10
$corr(\hat{F}, F)$		0.95	-	0.94	0.95	0.95	0.94	-	0.94	0.94	0.95
T = 200											
γ_2		1.01	0.01	0.96	1.00	1.05	0.95	-0.05	0.87	0.95	1.02
γ_1		1.00	-0.00	0.96	1.00	1.05	0.95	-0.05	0.85	0.96	1.02
ρ		0.90	-0.00	0.88	0.90	0.91	0.90	0.00	0.85	0.91	0.95
d		-0.01	-0.01	-0.08	-0.00	0.06	0.03	0.03	-0.04	0.04	0.12
σ_x		0.99	-0.01	0.94	0.98	1.04	1.03	0.03	0.97	1.05	1.11
σ_y		0.99	-0.01	0.93	0.99	1.05	1.05	0.05	0.95	1.09	1.18
$corr(\hat{F}, F)$		0.95	-	0.94	0.95	0.95	0.94	-	0.94	0.94	0.95
T = 100											
γ_2		0.99	-0.01	0.94	0.99	1.06	0.92	-0.08	0.79	0.90	1.02
γ_1		0.99	-0.01	0.94	0.98	1.05	0.89	-0.11	0.77	0.86	1.00
ρ		0.89	-0.01	0.87	0.90	0.91	0.89	-0.01	0.81	0.92	0.99
d		-0.02	-0.02	-0.11	-0.01	0.09	0.05	0.05	-0.16	0.09	0.23
σ_x		0.98	-0.02	0.91	0.99	1.05	1.03	0.03	0.89	1.09	1.20
σ_y		0.98	-0.02	0.90	0.97	1.08	1.13	0.13	0.99	1.19	1.29
$corr(\hat{F}, F)$		0.94	-	0.93	0.94	0.95	0.93	-	0.92	0.93	0.94

This table reports mean, bias, and 25%, 50%, 75% quantiles of the distribution of the ML (left) and Indirect Inference (II, right) estimators in 1000 MC replications. The data generating process is DGP1 in Section 5.1, corresponding to a mixed frequency linear state space model with a single AR(1) latent factor, $m = 3$, and stock sampling of the low frequency variable. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $\rho = 0.9$, $d = 0$, $\sigma_y = \sigma_x = 1$. The simulated samples have size $T = 500$ (top), $T = 200$ (middle), $T = 100$ (bottom). The auxiliary model for the indirect inference estimator is a U-MIDAS regression for low frequency data with $\tilde{K}_x = \tilde{K}_y = 3$ and an AR(9) model for the high frequency data (see equation (3.9)), with the correlation between the errors of the two equations as a free auxiliary parameter. The Indirect Inference estimator uses a single long simulated sample of the structural model ($S = 1$ and $T^S = 50000$) and an identity weighting matrix. We also compute the mean and 25%, 50%, 75% quantiles of the sample correlation between the estimated and true factor paths. The estimated factor paths are obtained by Kalman filter with the ML estimate (left) and the reprojection method with the II estimate (right), using $T^{reproj} = 100000$.

Table 3: MC simulations for the two-factor linear Gaussian state space model

Indirect Inference		T = 500					T = 200					T = 100				
Coeff.		mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.
$\gamma_{2,1}$		0.17	-0.03	0.00	0.00	0.37	0.16	-0.04	0.00	0.00	0.33	0.18	-0.02	0.00	0.12	0.36
$\gamma_{2,2}$		0.98	-0.02	0.92	0.99	1.05	0.98	-0.02	0.87	0.98	1.10	1.00	-0.00	0.85	1.00	1.17
$\gamma_{1,1}$		0.99	-0.01	0.91	0.98	1.05	1.05	0.05	0.94	1.03	1.17	1.01	0.01	0.81	0.99	1.22
$\gamma_{1,2}$		0.21	0.01	0.00	0.30	0.37	0.22	0.02	0.00	0.26	0.38	0.15	-0.05	0.00	0.12	0.27
ρ		0.90	-0.00	0.87	0.90	0.92	0.88	-0.02	0.84	0.89	0.92	0.88	-0.02	0.85	0.89	0.93
d		0.01	0.01	-0.07	0.00	0.07	-0.01	-0.01	-0.10	0.01	0.10	-0.05	-0.05	-0.18	-0.02	0.10
σ_x		1.00	-0.00	0.93	1.01	1.05	1.00	0.00	0.92	1.01	1.09	0.94	-0.06	0.85	1.00	1.09
σ_y		0.93	-0.07	0.85	0.98	1.09	0.76	-0.24	0.52	0.83	1.07	0.81	-0.19	0.36	0.96	1.21
$corr(\hat{F}_1, F_1)$		0.78	-	0.76	0.78	0.80	0.76	-	0.74	0.76	0.79	0.74	-	0.72	0.76	0.78
$corr(\hat{F}_2, F_2)$		0.92	-	0.91	0.92	0.92	0.91	-	0.90	0.92	0.93	0.88	-	0.89	0.92	0.93

This table reports mean, bias, and 25%, 50%, 75% quantiles of the distribution of the Indirect Inference (II) estimator in 1000 MC replications. The data generating process is DGP2 in Section 5.1, corresponding to a mixed frequency linear state space model with two independent $AR(1)$ latent factors, $m = 3$, and stock sampling of the low frequency variable. The true values of the parameters are $\gamma_1 = (1.0, 0.2)'$, $\gamma_2 = (0.2, 1.0)'$, $\rho = 0.9$, $d = 0$, $\sigma_y = \sigma_x = 1$. The simulated samples have size $T = 500$ (left), $T = 200$ (middle), $T = 100$ (right). The auxiliary model for the indirect inference estimator is a U-MIDAS regression for low frequency data with $\tilde{K}_x = \tilde{K}_y = 3$ and an $AR(9)$ model for the high frequency data (see equation (3.9)), with the correlation between the errors of the two equations as a free auxiliary parameter. The II estimator uses a single long simulated sample of the structural model ($S = 1$ and $T^S = 50000$) and an identity weighting matrix. We also compute the mean and 25%, 50%, 75% quantiles of the sample correlation between the estimated and true paths for each factor. The estimated factor paths are obtained by the reprojection method with the II estimate, using $T^{reproj} = 100000$.

Table 4: MC simulations for the stochastic volatility model (sample size $T = 500$)

$\rho = 0.5$	Indirect Inference (Auxiliary model: U-MIDAS/AR-ARCH)					MLE (Monte Carlo EM)					
	Coeff.	mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.
γ_2		0.97	-0.03	0.88	0.97	1.05	0.89	-0.11	0.86	0.89	0.92
γ_1		0.95	-0.05	0.87	0.94	1.01	1.07	0.07	1.02	1.06	1.12
ρ		0.51	0.01	0.47	0.51	0.55	0.52	0.02	0.51	0.53	0.54
d		-0.01	-0.01	-0.27	-0.02	0.29	-0.04	-0.04	-0.28	0.00	0.21
σ_y		0.95	-0.05	0.89	1.00	1.09	0.83	-0.17	0.75	0.84	0.91
σ_x		1.05	0.05	0.94	1.07	1.18	0.93	-0.07	0.91	0.94	0.96
ρ_{SV}		0.95	0.00	0.93	0.96	0.97	0.94	-0.01	0.93	0.94	0.95
ν		0.25	-0.05	0.19	0.25	0.31	0.29	-0.01	0.28	0.29	0.29
$corr(\hat{F}, F)$		0.74	-	0.72	0.74	0.75	0.79	-	0.78	0.79	0.80
$corr(\hat{h}, h)$		0.55	-	0.51	0.55	0.59	0.74	-	0.71	0.74	0.77
Comp. time (min)		15.89	-	9.01	14.69	20.94	61.05	-	38.49	54.11	78.05
$\rho = 0.9$											
γ_2		0.96	-0.04	0.91	0.96	1.00	0.97	-0.03	0.96	0.97	0.98
γ_1		0.95	-0.05	0.90	0.96	1.01	1.00	-0.00	0.98	1.00	1.02
ρ		0.91	0.01	0.89	0.91	0.93	0.90	0.00	0.89	0.90	0.91
d		0.00	0.00	-0.26	-0.05	0.38	-0.05	-0.05	-0.34	0.00	0.22
σ_y		0.97	-0.03	0.90	0.99	1.07	0.88	-0.12	0.81	0.88	0.95
σ_x		1.08	0.08	1.02	1.08	1.16	0.86	-0.14	0.82	0.87	0.91
ρ_{SV}		0.94	-0.01	0.93	0.95	0.97	0.95	-0.00	0.94	0.95	0.96
ν		0.26	-0.04	0.19	0.25	0.32	0.29	-0.01	0.28	0.29	0.29
$corr(\hat{F}, F)$		0.93	-	0.92	0.93	0.93	0.95	-	0.95	0.95	0.96
$corr(\hat{h}, h)$		0.51	-	0.47	0.51	0.56	0.75	-	0.72	0.76	0.79
Comp. time (min)		13.87	-	9.05	12.32	16.86	60.27	-	35.67	54.86	81.67

This table reports mean, bias, and 25%, 50%, 75% quantiles of the distribution of the Indirect Inference (II, left) and ML (right) estimators in 200 MC replications. The data generating process is DGP3 in Section 5.1, corresponding to a mixed frequency stochastic volatility model with a single AR(1) latent factor in the mean, an AR(1) log SV process, $m = 3$, and stock sampling of the low frequency variable. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $d = 0$, $\sigma_y = \sigma_x = 1$, $\rho_{SV} = 0.95$, $\nu = 0.3$. The autoregressive coefficient of the factor in the mean is $\rho = 0.5$ in the upper panel and $\rho = 0.9$ in the lower panel. The simulated samples have size $T = 500$. The auxiliary model for the II estimator is a U-MIDAS regression for low frequency data with $\tilde{K}_x = \tilde{K}_y = 4$ and an $AR(9) - ARCH(10)$ model for the high frequency data (see equation (3.10)), with the correlation between the errors of the two equations as a free auxiliary parameter. The II estimator uses a single long simulated sample of the structural model ($S = 1$ and $T^S = 50000$) and an identity weighting matrix. The MLE is computed by Monte Carlo EM, using a particle forward-filtering backward-smoothing algorithm in the E step (see Appendix B for the detailed algorithm). We also compute the mean and 25%, 50%, 75% quantiles of the sample correlation between the estimated and true paths of the mean and volatility factors. The estimated factor paths are obtained by the reprojection method with the Indirect Inference estimate (left), using $T^{reproj} = 100000$, and by the average across the particles of the filtering algorithm with the ML estimate (right). Finally, we report the mean and 25%, 50%, 75% quantiles of the computational time (in minutes) for obtaining the parameter estimates and the filtered factor paths in a single simulation.

Table 5: MC simulations for the stochastic volatility model (sample size $T = 200$)

$\rho = 0.5$	Indirect Inference (Auxiliary model: U-MIDAS/AR-ARCH)					MLE (Monte Carlo EM)					
	Coeff.	mean	bias	25% q.	median	75% q.	mean	bias	25% q.	median	75% q.
γ_2	0.99	-0.01	0.81	0.96	1.14	0.90	-0.10	0.86	0.91	0.94	
γ_1	0.88	-0.12	0.75	0.86	0.96	1.07	0.07	1.01	1.07	1.13	
ρ	0.50	0.00	0.44	0.51	0.60	0.52	0.02	0.49	0.52	0.56	
d	-0.05	-0.05	-0.35	-0.06	0.19	-0.03	-0.03	-0.31	0.00	0.27	
σ_y	0.96	-0.04	0.90	1.06	1.17	0.81	-0.19	0.71	0.82	0.92	
σ_x	1.03	0.03	0.90	1.09	1.24	0.94	-0.06	0.92	0.95	0.98	
ρ_{SV}	0.94	-0.01	0.93	0.96	0.99	0.94	-0.01	0.92	0.94	0.96	
ν	0.23	-0.07	0.08	0.22	0.34	0.29	-0.01	0.28	0.29	0.29	
$corr(\hat{F}, F)$	0.72	-	0.69	0.73	0.76	0.79	-	0.77	0.79	0.81	
$corr(\hat{h}, h)$	0.54	-	0.48	0.55	0.61	0.73	-	0.67	0.74	0.79	
Comp. time (min)	18.23	-	12.17	15.41	22.88	24.76	-	13.11	21.84	33.52	
$\rho = 0.9$											
γ_2	0.92	-0.08	0.84	0.93	1.00	0.97	-0.03	0.96	0.98	0.99	
γ_1	0.89	-0.11	0.79	0.89	0.98	1.00	0.00	0.98	1.00	1.03	
ρ	0.91	0.01	0.87	0.92	0.95	0.90	-0.00	0.89	0.90	0.91	
d	0.04	0.04	-0.25	-0.04	0.48	-0.01	-0.01	-0.28	0.00	0.28	
σ_y	1.03	0.03	0.88	1.06	1.18	0.87	-0.13	0.78	0.88	0.98	
σ_x	1.13	0.13	0.99	1.14	1.25	0.89	-0.11	0.85	0.91	0.95	
ρ_{SV}	0.93	-0.02	0.92	0.96	0.99	0.94	-0.01	0.92	0.94	0.96	
ν	0.22	-0.08	0.10	0.21	0.30	0.29	-0.01	0.28	0.29	0.29	
$corr(\hat{F}, F)$	0.92	-	0.91	0.92	0.93	0.95	-	0.95	0.95	0.96	
$corr(\hat{h}, h)$	0.51	-	0.46	0.52	0.58	0.74	-	0.70	0.74	0.80	
Comp. time (min)	16.45	-	12.11	14.19	19.94	22.50	-	11.50	18.33	32.10	

This table reports mean, bias, and 25%, 50%, 75% quantiles of the distribution of the Indirect Inference (II, left) and ML (right) estimators in 200 MC replications. The data generating process is DGP3 in Section 5.1, corresponding to a mixed frequency stochastic volatility model with a single AR(1) latent factor in the mean, an AR(1) log SV process, $m = 3$, and stock sampling of the low frequency variable. The true values of the parameters are $\gamma_1 = \gamma_2 = 1$, $d = 0$, $\sigma_y = \sigma_x = 1$, $\rho_{SV} = 0.95$, $\nu = 0.3$. The autoregressive coefficient of the factor in the mean is $\rho = 0.5$ in the upper panel and $\rho = 0.9$ in the lower panel. The simulated samples have size $T = 200$. The auxiliary model for the II estimator is a U-MIDAS regression for low frequency data with $\tilde{K}_x = \tilde{K}_y = 4$ and an $AR(9) - ARCH(10)$ model for the high frequency data (see equation (3.10)), with the correlation between the errors of the two equations as a free auxiliary parameter. The II estimator uses a single long simulated sample of the structural model ($S = 1$ and $T^S = 50000$) and an identity weighting matrix. The MLE is computed by Monte Carlo EM, using a particle forward-filtering backward-smoothing algorithm in the E step (see Appendix B for the detailed algorithm). We also compute the mean and 25%, 50%, 75% quantiles of the sample correlation between the estimated and true paths of the mean and volatility factors. The estimated factor paths are obtained by the reprojection method with the II estimate (left), using $T^{reproj} = 100000$, and by the average across the particles of the filtering algorithm with the ML estimate (right). Finally, we report the mean and 25%, 50%, 75% quantiles of the computational time (in minutes) for obtaining the parameter estimates and the filtered factor paths in a single simulation.

Table 6: In-sample R^2 of GDP and HF indicator on latent factor

HF Indicator	Stochastic volatility (Indirect Inference)		Gaussian state space (Kalman filter)	
	$R^2(GDP)$	$R^2(\text{HF indicator})$	$R^2(GDP)$	$R^2(\text{HF indicator})$
(1) Industrial Production	0.74	0.48	0.82	0.50
(2) Industrial Production - Pulp/paper	0.69	0.11	0.80	0.11
(3) Business Climate - IFO	0.56	0.43	0.42	0.65
(4) Economic Sentiment Index	0.48	0.71	0.36	0.86
(5) PMI Composite	0.49	0.46	0.80	0.04

This table reports the R^2 for the regressions of both GDP and the monthly indicators on the filtered values of the latent factor F for different mixed frequency models. We estimate the mixed-frequency stochastic volatility model defined as DGP 3 in Section 5 and the linear Gaussian state space model defined as DGP 1 with different pairs of mixed frequency observables. In each model, GDP is the low frequency (quarterly) observable, and is treated as a flow sampled variable. The table reports results for 10 different models, which differ for the high frequency (monthly) observable and the presence/absence of stochastic volatility. Columns 2 and 3 (resp. 4 and 5) display the R^2 for the regression of the GDP and the HF observable on the filtered values of F obtained from the models with (resp. without) stochastic volatility. We estimate the SV models via Indirect Inference using the same auxiliary models as in the MC simulations of Section 5. The mean and volatility factors are filtered by reprojection. We estimate the Gaussian state space model by adapting the Kalman filter for periodic state space models proposed in Bai, Ghysels, and Wright (2013), see Section C. All models are estimated on the full data sample from 1991-Q1 to 2011-Q1.

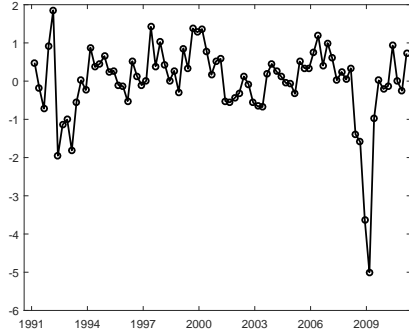
Table 7: Root Mean Squared Forecasting Error (RMSFE) for GDP

HF Indicator	Stochastic volatility (Indirect Inference)				Gaussian state space (Kalman filter)			
	Forecast horizon H (Quarters ahead)				Forecast horizon H (Quarters ahead)			
	1	2	3	4	1	2	3	4
(1) Industrial Production	0.79	1.00	1.12	1.16	0.70	1.02	1.06	1.06
(2) Industrial Production - Pulp/paper	0.83	1.30	1.00	1.05	0.80	1.02	1.03	1.02
(3) Business Climate - IFO	0.68	0.89	0.98	1.02	0.91	1.12	1.14	1.08
(4) Economic Sentiment Index	0.70	0.92	1.01	1.04	1.00	0.98	0.97	0.97
(5) PMI Composite	0.92	0.99	0.99	0.99	0.78	1.00	1.01	1.02

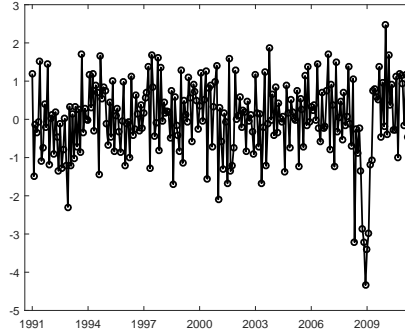
This table reports the Root Mean Squared Forecasting Error (RMSFE) for GDP in different mixed frequency models. The RMSFE is reported as the ratio to the RMSFE of the naive forecasting model assuming constant GDP growth rate. We consider the mixed-frequency stochastic volatility model defined as DGP 3 in Section 5 and the linear Gaussian state space model defined as DGP 1 with different pairs of mixed frequency observables. In each model, GDP is the low frequency (quarterly) observable, and is treated as a flow sampled variable. The table reports the forecasting results for 10 different models, which differ for the high frequency (monthly) observable and the presence/absence of stochastic volatility. To produce the forecasts, the models are estimated on the estimation window, and then used for prediction up to 4 quarters ahead of the estimation final date. The first estimation window is from 1991-Q1 to 2005-Q4, and is recursively expanded up to 2010-Q4. Columns 2 to 5 (resp. 6 and 9) display the RMSFE ratios at horizons $H = 1, 2, 3, 4$ quarters ahead for the models with (resp. without) stochastic volatility. We estimate the SV models via Indirect Inference using the same auxiliary models as in the MC simulations of Section 5. We estimate the Gaussian state space model by adapting the Kalman filter for periodic state space models proposed in Bai, Ghysels, and Wright (2013), see Section C.

FIGURES

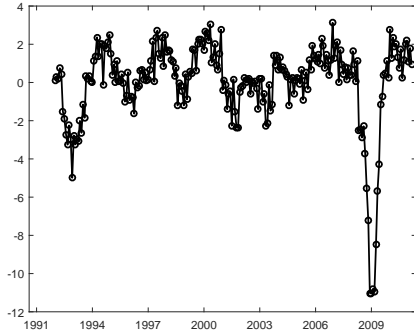
Figure 1: Time series of observables and estimated factors: stochastic volatility model estimated on GDP and IP.



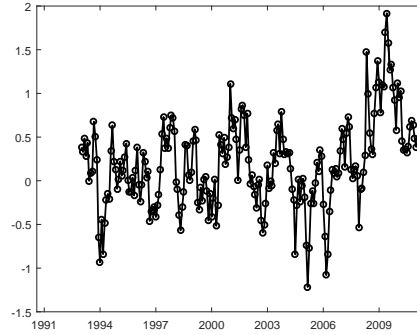
(a) Quarterly GDP.



(b) Monthly IP index.



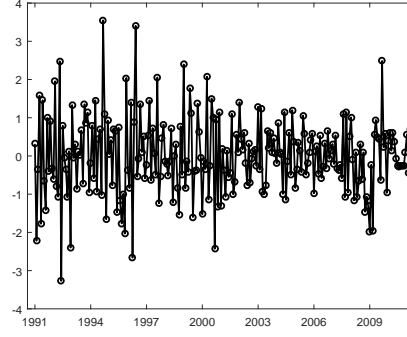
(c) Estimated common factor \hat{F} .



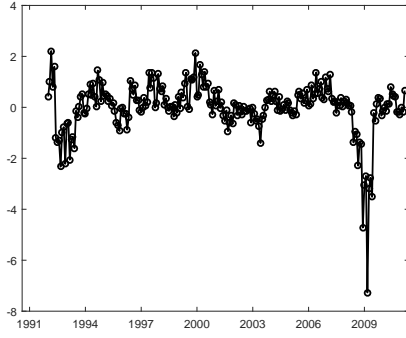
(d) Estimated SV factor \hat{h} .

Panels (a) and (b) display the time series of the (standardized) quarterly growth rate of European GDP, and the (standardized) monthly growth rate of aggregate European Industrial Production index. These series are used to estimate the mixed frequency state space model with stochastic volatility specified as DGP 3 in Section 5, with flow sampling of the low frequency variable. The sample is from 1991-Q1 to 2011-Q1. Panels (c) and (d) display the estimated mean and idiosyncratic volatility factors \hat{F} and \hat{h} . The parameters of the SV model are estimated via Indirect Inference, using the same auxiliary model as in the MC simulations of Section 5. The factors are filtered by reprojection.

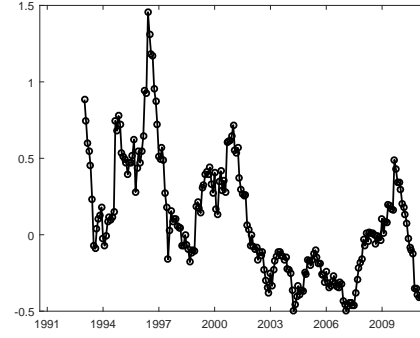
Figure 2: Time series of high frequency observable and estimated factors: stochastic volatility model estimated on GDP and IP - Pulp/Paper.



(a) Monthly IP - Pulp/Paper index.



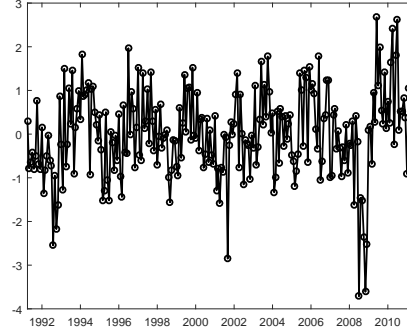
(b) Estimated common factor $\hat{F}_{t+j/3}$.



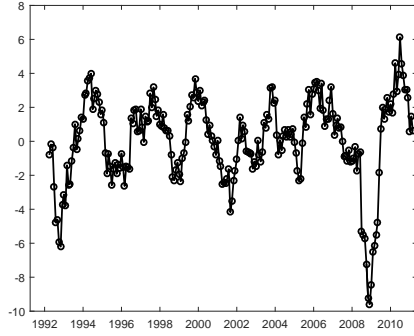
(c) Estimated SV factor $\hat{h}_{t+j/3}$.

Panel (a) displays the (standardized) monthly growth rate of the European Industrial Production index for “Pulp and Paper sector”. These series are used to estimate the mixed frequency state space model with stochastic volatility specified as DGP 3 in Section 5, with flow sampling of the low frequency variable (European GDP, represented in Figure 1 (a)). The sample is from 1991-Q1 to 2011-Q1. Panels (b) and (c) display the estimated mean and idiosyncratic volatility factors \hat{F} and \hat{h} . The parameters of the SV model are estimated via Indirect Inference, using the same auxiliary model as in the MC simulations of Section 5. The factors are filtered by reprojection.

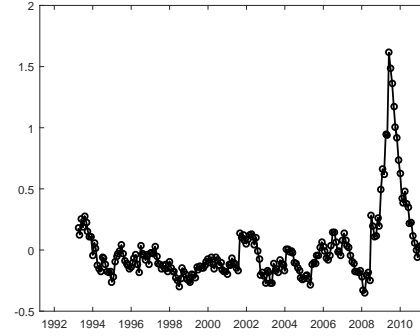
Figure 3: Time series of high frequency observables and estimated factors: stochastic volatility model estimated on GDP and IFO.



(a) Monthly IFO index.



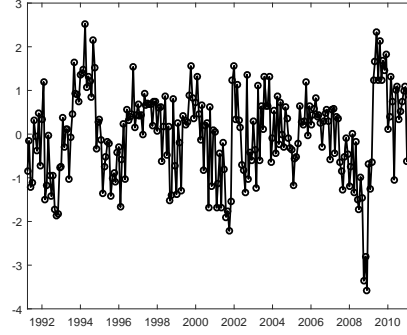
(b) Estimated common factor $\hat{F}_{t+j/3}$.



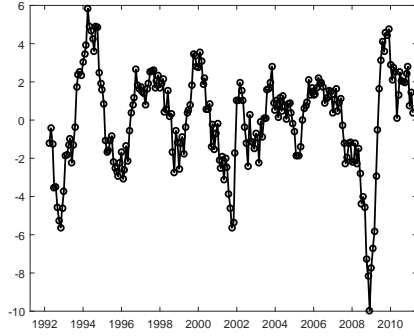
(c) Estimated SV factor $\hat{h}_{t+j/3}$.

Panel (a) displays the time series of the monthly (standardized first difference of) Germany IFO Business Climate index. These series are used to estimate the mixed frequency state space model with stochastic volatility specified as DGP 3 in Section 5, with flow sampling of the low frequency variable (European GDP, represented in Figure 1 (a)). The sample is from 1991-Q1 to 2011-Q1. Panels (b) and (c) display the estimated mean and idiosyncratic volatility factors \hat{F} and \hat{h} . The parameters of the SV model are estimated via Indirect Inference, using the same auxiliary model as in the MC simulations of Section 5. The factors are filtered by reprojection.

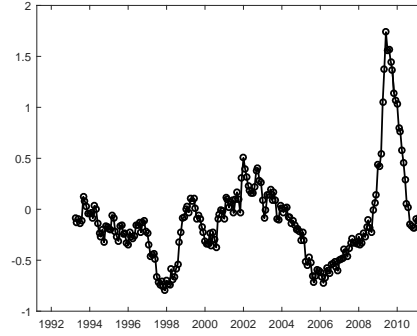
Figure 4: Time series of high frequency observable and estimated factors: stochastic volatility model estimated on GDP and ESI.



(a) Monthly ESI index.



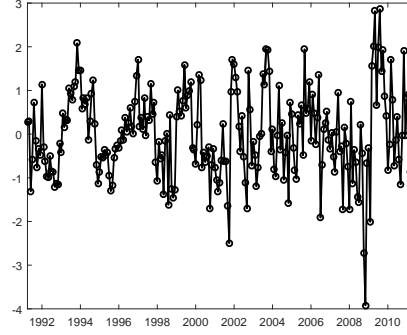
(b) Estimated common factor $\hat{F}_{t+j/3}$.



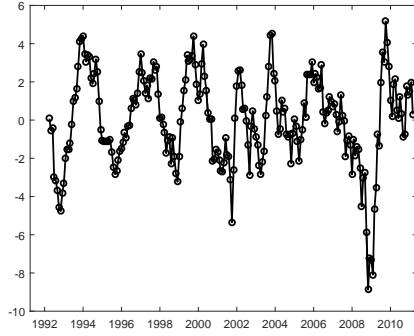
(c) Estimated SV factor $\hat{h}_{t+j/3}$.

Panel (a) displays the time series of the (standardized first difference of) monthly Euro-area Economic Sentiment Index. These series are used to estimate the mixed frequency state space model with stochastic volatility specified as DGP 3 in Section 5, with flow sampling of the low frequency variable (European GDP, represented in Figure 1 (a)). The sample is from 1991-Q1 to 2011-Q1. Panels (b) and (c) display the estimated mean and idiosyncratic volatility factors \hat{F} and \hat{h} . The parameters of the SV model are estimated via Indirect Inference, using the same auxiliary model as in the MC simulations of Section 5. The factors are filtered by reprojection.

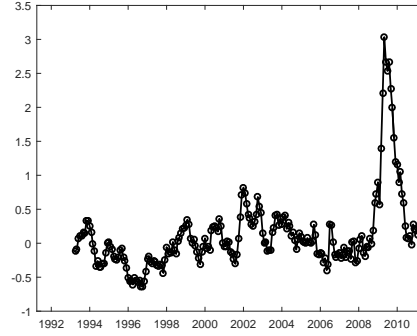
Figure 5: Time series of high frequency observable and estimated factors: stochastic volatility model estimated on GDP and PMI.



(a) Monthly PMI index.



(b) Estimated common factor $\hat{F}_{t+j/3}$.



(c) Estimated SV factor $\hat{h}_{t+j/3}$.

Panel (a) displays the time series of the (standardized first difference of) monthly Euro-area Composite Purchasing Manager Index. These series are used to estimate the mixed frequency state space model with stochastic volatility specified as DGP 3 in Section 5, with flow sampling of the low frequency variable (European GDP, represented in Figure 1 (a)). The sample is from 1991-Q1 to 2011-Q1. Panels (b) and (c) display the estimated mean and idiosyncratic volatility factors \hat{F} and \hat{h} . The parameters of the SV model are estimated via Indirect Inference, using the same auxiliary model as in the MC simulations of Section 5. The factors are filtered by reprojection.

Appendix A: Mixed Frequency Linear State Space Models

In this Appendix we summarize some results from Bai, Ghysels, and Wright (2013) concerning linear state space models with mixed frequency data. These results are useful to obtain the binding function linking our structural state space model and the auxiliary MIDAS regressions when the structural model does not feature SV (see Section 3.1). They also provide the Kalman filter algorithm for ML estimation of the structural model without SV used in the MC simulations (see Section 5).

A.1 Model setup

The linear state space model presented in Section 3.1 can be summarized as follows. The latent factor F follows a $VAR(p)$ process:

$$F_{t+j/m} = \sum_{l=1}^p \Phi_l F_{t+(j-l)/m} + \eta_{t+j/m} \quad \forall t = 1, \dots, T, \quad j = 0, \dots, m-1. \quad (\text{A.1})$$

The low frequency data is related to factors as follows:

$$y_{t+j/m}^* = \gamma_1' F_{t+j/m} + u_{1,t+j/m} \quad \forall t, \quad j = 0, \dots, m-1, \quad (\text{A.2})$$

with $u_{1,t+j/m}$ having an $AR(k)$ representation:

$$d_1(L^{1/m})u_{1,t+j/m} = \varepsilon_{1,t+j/m}, \quad d_1(L^{1/m}) \equiv 1 - d_{11}L^{1/m} - \dots - d_{k1}L^{k/m}, \quad (\text{A.3})$$

and the lag operator $L^{1/m}$ applying to high-frequency data, i.e. $L^{1/m}u_t \equiv u_{t-1/m}$. The observed low-frequency process y relates to the latent process y^* via a linear aggregation scheme:

$$y_{t+j/m}^c = \Psi_j y_{t+(j-1)/m}^c + \lambda_j y_{t+j/m}^* \quad (\text{A.4})$$

where y_t is equal to y_t^c for integer t , and is not observed otherwise. The high frequency process $x_{t+j/m}$ relates to the factors as follows:

$$x_{t+j/m} = \gamma_2' F_{t+j/m} + u_{2,t+j/m} \quad \forall t, \quad j = 0, \dots, m-1, \quad (\text{A.5})$$

where:

$$d_2(L^{1/m})u_{2,t+j/m} = \varepsilon_{2,t+j/m}, \quad d_2(L^{1/m}) \equiv 1 - d_{12}L^{1/m} - \dots - d_{k2}L^{k/m}. \quad (\text{A.6})$$

This model corresponds to a restricted version of the specification in Assumptions 2.1 and 2.2 with $\rho_{SV} = 1$ and $\nu_2 = 0$.

A.2 State space representation and Kalman filter

The above equations yield a periodic state space model with measurement equation:

$$Y_t^j = Z_j \alpha_{t+j/m} \quad \begin{cases} Y_t^j = (y_t, x_t)' & j = 0 \\ Y_t^j = x_{t+j/m} & 0 < j \leq m-1 \end{cases}, \quad (\text{A.7})$$

where

$$Z_0 = \begin{bmatrix} \gamma'_1 & O_{2 \times n_f(p-1)} & I_2 & O_{2 \times 2(k-1)} \\ \gamma'_2 & & & \end{bmatrix}$$

$$Z_j = \begin{bmatrix} \gamma'_2 & O_{1 \times n_f(p-1)} & 1 & O_{1 \times 2(k-1)} \end{bmatrix}$$

for $0 < j \leq m-1$ and state vector

$$\alpha_{t+j/m} = \left(F'_{t+j/m}, \dots, F'_{t+(j-p+1)/m}, u'_{t+j/m}, \dots, u'_{t+(j-k+1)/m} \right)'$$

where $u_{t+j/m} = (u_{1,t+j/m}, u_{2,t+j/m})'$.

The transition equation is:

$$\alpha_{t+j/m} = R \alpha_{t+(j-1)/m} + Q \zeta_{t+j/m} \quad (\text{A.8})$$

where

$$R = \begin{bmatrix} \Phi_1 \dots \Phi_{p-1} & \Phi_p & O_{n_f \times 2(k-1)} & O_{n_f \times 2} \\ I_{(p-1)n_f} & O_{(p-1)n_f \times n_f} & O_{(p-1)n_f \times 2(k-1)} & O_{(p-1)n_f \times n} \\ O_{2 \times (p-1)n_f} & O_{2 \times n_f} & D_1 \dots D_{k-1} & D_k \\ O_{2(k-1) \times (p-1)n_f} & O_{2(k-1) \times n_f} & I_{2(k-1)} & O_{2(k-1) \times 2} \end{bmatrix}$$

$$Q = \begin{bmatrix} I_{n_f} & O_{n_f \times 2} \\ O_{(p-1)n_f \times n_f} & O_{(p-1)n_f \times 2} \\ O_{2 \times n_f} & I_2 \\ O_{2(k-1) \times n_f} & O_{2(k-1) \times 2} \end{bmatrix}$$

$D_i = \text{diag}(d_{li}, l = 1, 2)$ and $\zeta_{t+j/m} = (\eta'_{t+j/m}, \varepsilon_{1,t+j/m}, \varepsilon_{2,t+j/m})'$. Let Σ_ζ denote the variance-covariance matrix of $\zeta_{t+j/m}$.

The above state space model is periodic as it cycles to the data release pattern that repeats itself every m periods. Such systems have a (periodic) steady state (see e.g. Assimakis and Adam (2009)). If we let $P_{j|j-1}$ denote the steady state covariance matrix of $\alpha_{t+j/m|t+(j-1)/m}$, then the equations:

$$\begin{aligned} P_{j+1|j} &= Q \Sigma_\zeta Q' + R P_{j|j-1} R' - R P_{j|j-1} Z'_j [Z_j P_{j|j-1} Z'_j]^{-1} Z_j P_{j|j-1} R' & j = 0, \dots, m-2 \\ P_{0|-1} &= Q \Sigma_\zeta Q' + R P_{2|1} R' - R P_{2|1} Z'_j [Z_j P_{2|1} Z'_j]^{-1} Z_j P_{2|1} R' & j = m-1 \end{aligned} \quad (\text{A.9})$$

must be satisfied and $P_{j|j-1} = P_{j+m|j+m-1}, \forall j$. The periodic steady state Kalman gain is therefore:

$$K_{j|j-1} = P_{j|j-1} Z'_j [Z_j P_{j|j-1} Z'_j]^{-1} \quad (\text{A.10})$$

with $K_{j|j-1} \equiv K_{j+m|j-1+m}, \forall j$. Let us define the information set $I_{t+j/m}^M$ as the linear space generated by $\{Y_{t+(j-k)/m}^j | k \geq 0\}$. When we define the extraction of the state vector as:

$$\hat{\alpha}_{(t+j/m)|(t+j/m)} = E[\alpha_{t+j/m} | I_{t+j/m}^M] \quad (\text{A.11})$$

the filtered states are:

$$\hat{\alpha}_{(t+j/m)|(t+j/m)} = A_{j|j-1}\hat{\alpha}_{t+(j-1)/m|t+(j-1)/m} + K_{j|j-1}Y_t^j \quad (\text{A.12})$$

where $A_{j|j-1} = R - K_{j|j-1}Z_jR$ and $Y_t^m = Y_{t+1}^0$.

Suppose we are interested in predicting at low-frequency intervals only, namely $\hat{\alpha}_{(t+k)|t}$, for k integer valued, using all available low and high-frequency data. First we note that:

$$\hat{\alpha}_{(t+k)|(t+k)} = [\tilde{A}_1^m]^k \hat{\alpha}_{t|t} + \sum_{i=1}^m \sum_{j=1}^k [\tilde{A}_1^m]^{k-j} \tilde{A}_{i+1}^m K_{i|i-1} Y_{t+j-1}^i \quad (\text{A.13})$$

where

$$\tilde{A}_j^i = \begin{cases} A_{i|i-1} A_{i-1|i-2} \dots A_{j|j-1} & i \geq j \\ I & i < j \end{cases}$$

Expression (A.13) can be obtained via straightforward algebra - see Assimakis and Adam (2009). Given Assumption 2.1, all the eigenvalues of $A_{j|j-1}$, $j = 1, \dots, m-1$, are inside the unit circle, as are the eigenvalues of the product matrices $\{\tilde{A}_j^i\}$ (see again Assimakis and Adam (2009)). This implies that we can rewrite (A.13) as:

$$\begin{aligned} \hat{\alpha}_{t|t} &= \sum_{j=0}^{\infty} \sum_{i=1}^m [\tilde{A}_1^m]^j \tilde{A}_{i+1}^m K_{i|i-1} Y_{t-j}^i \\ &= \sum_{j=0}^{\infty} [\tilde{A}_1^m]^j K_{m|m-1} \begin{pmatrix} y_{t-j} \\ x_{t-j} \end{pmatrix} + \sum_{j=0}^{\infty} \sum_{i=1}^{m-1} [\tilde{A}_1^m]^j \tilde{A}_{i+1}^m K_{i|i-1} x_{t-1-j+i/m} \end{aligned} \quad (\text{A.14})$$

from which forecasts can easily be constructed as $E_t[y_{t+h}] = Z_{0,1} R^{mh} \hat{\alpha}_{t|t}$, where $Z_{0,1}$ denotes the first row of the matrix Z_0 . When factor F is scalar with autoregressive coefficient ρ , and $m = 3$, the latter equation yields equation (3.1) in Section 3.1.

A.3 ML estimation

To proceed to maximum likelihood estimation, let $\theta \in \Theta$ be the parameter vector governing the parameters of the state space model, i.e. $\theta = ((\gamma_i)_{i=1}^2, (\Psi_i)_{i=1}^p, (D_i)_{i=1}^k, \Sigma_\zeta)$ (accounting for identification constraints). Consider the vector Y_t^j defined for $j = 0, \dots, m-1$, in equation (A.7) and the information set $I_{t+j/m}^M$ in equation (A.11). Then:

$$Y_{t+(j+1)/m}^j | I_{t+j/m}^M; \theta \sim \mathcal{N}(\mu_{t+(j+1)/m}(\theta), \Sigma_{t+(j+1)/m}(\theta)) \quad (\text{A.15})$$

with $\mu_{t+(j+1)/m}(\theta) \equiv Z_{j+1}(\theta) \hat{\alpha}_{t+(j+1)/m|t+j/m}(\theta)$ and

$$\Sigma_{t+(j+1)/m}(\theta) \equiv Z_{j+1}(\theta)' P_{t+(j+1)/m|t+j/m}(\theta) Z_{j+1}(\theta) + Q(\theta).$$

The value of the log likelihood for a sample of size Tm is then:

$$\begin{aligned} \sum_{t=1}^T \sum_{j=0}^{m-1} \log \ell(Y_{t+(j+1)/m}^j | I_{t+j/m}^M; \theta) &= -\frac{Tm}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log |\Sigma_{t+(j+1)/m}(\theta)| \\ &\quad - \frac{1}{2} \sum_{t=1}^T \sum_{j=0}^{m-1} (Y_{t+(j+1)/m}^j - \mu_{t+(j+1)/m})' (\Sigma_{t+(j+1)/m}(\theta))^{-1} (Y_{t+(j+1)/m}^j - \mu_{t+(j+1)/m}(\theta)) \end{aligned} \quad (\text{A.16})$$

We denote the estimator that maximizes this log likelihood function by $\hat{\theta}_{Tm}^{ML}$. Standard regularity conditions imply that as $T \rightarrow \infty$:

$$\sqrt{Tm}(\hat{\theta}_{Tm}^{ML} - \theta_0) \rightarrow_d \mathcal{N}(0, V^{ML}), \quad (\text{A.17})$$

where θ_0 denotes the true parameter value.

Appendix B: Estimation of the mixed-frequency SV model by Monte Carlo EM algorithm

In this Appendix we describe a Monte Carlo Expectation Maximization (EM) algorithm for estimation of the state space model with mixed frequency data and stochastic volatility (see Section 3.5). In this algorithm, the smoothing distribution of the latent factors necessary in the Expectation step is obtained using a Forward Filtering-Backward Smoothing simulation-based procedure.

B.1 Model Setup

In this appendix we only consider models with unidimensional observables y_t and $x_{t+j/m}$, and unidimensional latent factors $F_{t+j/m}$ and $h_{t+j/m}$. The generalization to multivariate observables and latent factors is relatively straightforward, at the expense of a more involved notation. We consider a model with autocorrelated innovations $u_{y,t+j/m}$ and stock sampled LF variables $y_{t+j/m}^*$:

$$y_{t+j/m}^* = \gamma_1 F_{t+j/m} + u_{y,t+j/m}, \quad (\text{B.1})$$

$$x_{t+j/m} = \gamma_2 F_{t+j/m} + u_{x,t+j/m}, \quad (\text{B.2})$$

$$F_{t+j/m} = \rho F_{t+(j-1)/m} + \eta_{t+j/m}, \quad (\text{B.3})$$

$$h_{t+j/m} = \rho_{SV} h_{t+(j-1)/m} + \nu \xi_{t+j/m}, \quad (\text{B.4})$$

$$u_{y,t+j/m} = d u_{y,t+(j-1)/m} + \sigma_y \varepsilon_{y,t+j/m}, \quad (\text{B.5})$$

$$u_{x,t+j/m} = \sigma_x \exp\left\{\frac{1}{2} h_{t+j/m}\right\} \varepsilon_{x,t+j/m}, \quad (\text{B.6})$$

$$(\eta_{t+j/m}, \xi_{t+j/m}, \varepsilon_{y,t+j/m}, \varepsilon_{x,t+j/m})' \sim i.i.\mathcal{N}(0, I_4), \quad (\text{B.7})$$

$$y_{t+j/m}^* \quad \text{is stock-sampled at } j = 0. \quad (\text{B.8})$$

We focus on the setting with $m = 3$ as in the Monte-Carlo analysis of Section 5.

In Section B.2 we derive the state space representation of the SV model in low frequency. In Section B.3 we describe the E-step and the M-step of the EM algorithm. In Section B.4 we provide the simulation-based procedure to obtain the smoothing distribution of the latent factor process required in the E-step of the EM algorithm. Throughout this Appendix, $\ell(\cdot)$ denotes the (conditional) density of the indicated random variables.

B.2 State space representation

We derive a state space representation of model (B.1)-(B.8) in low frequency. For this purpose, we define the vector of stacked measurements Y_t and the vector of stacked latent factors f_t as follows:

$$Y_t := (y_t, x_t, x_{t-1/3}, x_{t-2/3})',$$

$$f_t := \begin{bmatrix} \tilde{F}_t \\ \tilde{h}_t \end{bmatrix}, \quad \tilde{F}_t := \begin{bmatrix} F_t \\ F_{t-1/3} \\ F_{t-2/3} \end{bmatrix}, \quad \tilde{h}_t := \begin{bmatrix} h_t \\ h_{t-1/3} \\ h_{t-2/3} \end{bmatrix}.$$

B.2.1 Measurement density

Let us first derive the distribution of Y_t given $\underline{Y_{t-1}}$ and $\underline{f_t}$. From equations (B.1)-(B.8) we get:

$$Y_t = \Gamma \tilde{F}_t + u_t, \quad (\text{B.9})$$

where

$$u_t := (u_{y,t}, u_{x,t}, u_{x,t-1/3}, u_{x,t-2/3})', \quad \Gamma := \begin{bmatrix} \gamma_1 & 0 & 0 \\ \gamma_2 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & \gamma_2 \end{bmatrix}.$$

To derive the dynamics of innovation u_t , we use that equation (B.5) and backward iteration imply $u_{y,t} = d^3 u_{y,t-1} + \sigma_y (\varepsilon_{y,t} + d \varepsilon_{y,t-1/3} + d^2 \varepsilon_{y,t-2/3})$. This equation can be written as:

$$u_{y,t} = d^3 u_{y,t-1} + \sigma_y \sqrt{1 + d^2 + d^4} \varepsilon_{y,t}^*, \quad \varepsilon_{y,t}^* \sim i.i.\mathcal{N}(0, 1), \quad t = 1, 2, \dots, T,$$

where $(\varepsilon_{y,t}^*)$ is independent from $(\varepsilon_{x,t-j/3})$, $(\eta_{t-j/3})$ and $(\xi_{t-j/3})$. Thus, innovation process (u_t) is such that:

$$u_t = A u_{t-1} + B_t \tilde{\varepsilon}_t^*, \quad (\text{B.10})$$

where

$$\tilde{\varepsilon}_t^* = \begin{bmatrix} \varepsilon_{y,t}^* \\ \varepsilon_{x,t} \\ \varepsilon_{x,t-1/3} \\ \varepsilon_{x,t-2/3} \end{bmatrix}, \quad A = \begin{bmatrix} d^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$B_t = \begin{bmatrix} \sigma_y \sqrt{1 + d^2 + d^4} & 0 & 0 & 0 \\ 0 & \sigma_x \exp\{\frac{1}{2} h_t\} & 0 & 0 \\ 0 & 0 & \sigma_x \exp\{\frac{1}{2} h_{t-1/3}\} & 0 \\ 0 & 0 & 0 & \sigma_x \exp\{\frac{1}{2} h_{t-2/3}\} \end{bmatrix}.$$

Equations (B.9) and (B.10) imply:

$$Y_t - A Y_{t-1} = \Gamma \tilde{F}_t - A \Gamma \tilde{F}_{t-1} + B_t \tilde{\varepsilon}_t^*,$$

and thus:

$$Y_t = \begin{bmatrix} y_t \\ x_t \\ x_{t-1/3} \\ x_{t-2/3} \end{bmatrix} = \begin{bmatrix} d^3 y_{t-1} + \gamma_1 (F_t - d^3 F_{t-1}) \\ \gamma_2 F_t \\ \gamma_2 F_{t-1/3} \\ \gamma_2 F_{t-2/3} \end{bmatrix} + B_t \tilde{\varepsilon}_t^*.$$

From the last equation we get the measurement distribution:

$$Y_t \mid \underline{Y_{t-1}}, \underline{f_t} \sim \mathcal{N} \left(\begin{bmatrix} d^3 y_{t-1} + \gamma_1 (F_t - d^3 F_{t-1}) \\ \gamma_2 F_t \\ \gamma_2 F_{t-1/3} \\ \gamma_2 F_{t-2/3} \end{bmatrix}, B_t^2 \right), \quad (\text{B.11})$$

and the measurement density:

$$\begin{aligned}
\ell(Y_t|\underline{Y}_{t-1}, \underline{f}_t; \theta) &= \frac{1}{\sqrt{(2\pi)^4 \sigma_y^2 (1 + d^2 + d^4) (\sigma_x^2)^3 \exp\{h_t + h_{t-1/3} + h_{t-2/3}\}}} \\
&\times \exp \left\{ - \frac{[y_t - d^3 y_{t-1} - \gamma_1 (F_t - d^3 F_{t-1})]^2}{2(1 + d^2 + d^4) \sigma_y^2} \right. \\
&\quad \left. - \frac{(x_t - \gamma_2 F_t)^2}{2\sigma_x^2 \exp\{h_t\}} - \frac{(x_{t-1/3} - \gamma_2 F_{t-1/3})^2}{2\sigma_x^2 \exp\{h_{t-1/3}\}} - \frac{(x_{t-2/3} - \gamma_2 F_{t-2/3})^2}{2\sigma_x^2 \exp\{h_{t-2/3}\}} \right\} \\
&=: h(Y_t|Y_{t-1}, \underline{f}_t; \theta).
\end{aligned} \tag{B.12}$$

The measurement density depends on the past measurement Y_{t-1} , and on the current and past factor values f_t, f_{t-1} .

B.2.2 Transition density

Let us now derive the distribution of f_t given Y_{t-1} and f_{t-1} . From equations (B.3)-(B.4) and being $(\eta_{t+j/3}), (\xi_{t+j/3})$ independent Gaussian White Noise processes, we have:

$$\ell(f_t|\underline{Y}_{t-1}, \underline{f}_{t-1}; \theta) = \ell(f_t|f_{t-1}; \theta) = \ell(\tilde{F}_t|\tilde{F}_{t-1}; \theta) \ell(\tilde{h}_t|\tilde{h}_{t-1}; \theta).$$

Thus, process (f_t) is exogenous and first-order Markov, with transition density:

$$\begin{aligned}
g(f_t|f_{t-1}; \theta) &= g(\tilde{F}_t|\tilde{F}_{t-1}; \theta) g(\tilde{h}_t|\tilde{h}_{t-1}; \theta) \\
&= g(F_t|F_{t-1/3}; \theta) g(F_{t-1/3}|F_{t-2/3}; \theta) g(F_{t-2/3}|F_{t-1}; \theta) \\
&\quad \times g(h_t|h_{t-1/3}; \theta) g(h_{t-1/3}|h_{t-2/3}; \theta) g(h_{t-2/3}|h_{t-1}; \theta),
\end{aligned}$$

where:

$$g(F_{t-j/3}|F_{t-(j+1)/3}; \theta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ - \frac{(F_{t-j/3} - \rho F_{t-(j+1)/3})^2}{2} \right\}, \tag{B.13}$$

$$g(h_{t-j/3}|h_{t-(j+1)/3}; \theta) = \frac{1}{\sqrt{2\pi\nu^2}} \exp \left\{ - \frac{(h_{t-j/3} - \rho_{SV} h_{t-(j+1)/3})^2}{2\nu^2} \right\}, \tag{B.14}$$

for $j = 0, 1, 2$.

B.2.3 The likelihood function

The density of (Y_T, f_T) , conditioning on Y_0 and f_0 , is:

$$\ell(\underline{Y}_T, \underline{f}_T; \theta) = \prod_{t=1}^T \ell(Y_t|\underline{Y}_{t-1}, \underline{f}_t; \theta) \ell(f_t|\underline{Y}_{t-1}, \underline{f}_{t-1}; \theta) = \prod_{t=1}^T h(Y_t|Y_{t-1}, \underline{f}_t; \theta) g(f_t|f_{t-1}; \theta).$$

The likelihood function $\ell(\underline{Y}_T; \theta)$, conditioning on y_0 and f_0 , is obtained by integrating out the path of the unobservable factor:

$$\begin{aligned}\ell(\underline{Y}_T; \theta) &= \int \ell(\underline{Y}_T, \underline{f}_T; \theta) d\underline{f}_T \\ &= \int \dots \int \prod_{t=1}^T \{h(Y_t | Y_{t-1}, \underline{f}_t; \theta) g(f_t | f_{t-1}; \theta)\} \prod_{t=1}^T df_t.\end{aligned}$$

The large-dimensional integral with respect to the factor path makes this expression of the likelihood function intractable for the computation of the Maximum Likelihood (ML) estimate. The EM algorithm defined in the next section relies instead on the so-called *complete-observation log-likelihood function*, i.e., the log-density function of both the observable and unobservable variables:

$$\begin{aligned}\mathcal{L}^*(\theta) &:= \log \ell(\underline{Y}_T, \underline{f}_T; \theta) \\ &= \sum_{t=1}^T \log h(Y_t | Y_{t-1}, \underline{f}_t; \theta) + \sum_{t=1}^T \log g(f_t | f_{t-1}; \theta) \\ &= \sum_{t=1}^T \log h(Y_t | Y_{t-1}, \underline{f}_t; \theta) \\ &\quad + \sum_{t=1}^T \sum_{j=0}^2 [\log g(F_{t-j/3} | F_{t-(j+1)/3}; \theta) + \log g(h_{t-j/3} | h_{t-(j+1)/3}; \theta)].\end{aligned}\tag{B.15}$$

Substituting equations (B.12), (B.13) and (B.14) into equation (B.15) we get:

$$\begin{aligned}\mathcal{L}^*(\theta) &= -\frac{1}{2} \left(T \log(1 + d^2 + d^4) + T \log \sigma_y^2 + 3T \log \sigma_x^2 + 3T \log \nu^2 \right. \\ &\quad + \sum_{t=1}^T \frac{[y_t - d^3 y_{t-1} - \gamma_1(F_t - d^3 F_{t-1})]^2}{(1 + d^2 + d^4) \sigma_y^2} \\ &\quad + \sum_{t=1}^T \sum_{j=0}^2 \left\{ h_{t-j/3} + \frac{(x_{t-j/3} - \gamma_2 F_{t-j/3})^2}{\sigma_x^2 \exp\{h_{t-j/3}\}} + (F_{t-j/3} - \rho F_{t-(j+1)/3})^2 \right. \\ &\quad \left. \left. + \frac{1}{\nu^2} (h_{t-j} - \rho_{SV} h_{t-(j+1)/3})^2 \right\} \right),\end{aligned}\tag{B.16}$$

up to an additive constant.

B.3 The EM algorithm

The Expectation-Maximization (EM) algorithm is an iterative procedure to compute numerically the ML estimate in a model with unobservable variables (Dempster, Laird, and Rubin (1977)). Let $\hat{\theta}^{(i)} \equiv \hat{\theta}_{T_m}^{EM, (i)}$ be the estimate of parameter θ at iteration i of the EM algorithm. The update $i \rightarrow i + 1$ consists of two steps:

1. *Expectation (E) step.* Compute function $Q(\theta|\tilde{\theta})$, with $\tilde{\theta} = \hat{\theta}^{(i)}$, where:

$$\begin{aligned} Q(\theta|\tilde{\theta}) &:= E_{\tilde{\theta}} [\mathcal{L}^*(\theta)|\underline{Y}_T] \\ &= \sum_{t=1}^T E_{\tilde{\theta}} [h(Y_t|Y_{t-1}, \underline{f}_t; \theta)|\underline{Y}_T] \\ &\quad + \sum_{t=1}^T \sum_{j=0}^2 E_{\tilde{\theta}} [\log g(F_{t-j/3}|F_{t-(j+1)/3}; \theta) + \log g(h_{t-j/3}|h_{t-(j+1)/3}; \theta)|\underline{Y}_T], \end{aligned}$$

and $E_{\tilde{\theta}} [\cdot|\underline{Y}_T]$ denotes the expectation w.r.t. the conditional distribution of \underline{f}_T given \underline{Y}_T for parameter value $\tilde{\theta}$.

2. *Maximization (M) step.* Compute the estimate for iteration $i + 1$ as:

$$\hat{\theta}^{(i+1)} := \arg \max_{\theta \in \Theta} Q(\theta|\hat{\theta}^{(i)}).$$

The iteration is performed until numerical convergence of the estimate is achieved.

We detail below the E-step and the M-step of the EM algorithm for the SV model with mixed frequency.

B.3.1 The E-step

Let us compute explicitly $Q(\theta|\tilde{\theta})$, with $\tilde{\theta} = \hat{\theta}^{(i)}$, for model (B.1)-(B.8). From (B.16), we have:

$$\begin{aligned} Q(\theta|\tilde{\theta}) &:= E_{\tilde{\theta}} [\mathcal{L}^*(\theta)|\underline{Y}_T] \\ &= -\frac{1}{2} \left(T \log(1 + d^2 + d^4) + T \log \sigma_y^2 + 3T \log \sigma_x^2 + 3T \log \nu^2 + \sum_{t=1}^T \sum_{j=0}^2 E_{\tilde{\theta}} [h_{t-j/3}|\underline{Y}_T] \right. \\ &\quad + \frac{1}{(1 + d^2 + d^4)\sigma_y^2} \sum_{t=1}^T E_{\tilde{\theta}} [(y_t - d^3 y_{t-1} - \gamma_1(F_t - d^3 F_{t-1}))^2|\underline{Y}_T] \\ &\quad + \frac{1}{\sigma_x^2} \sum_{t=1}^T \sum_{j=0}^2 \left\{ E_{\tilde{\theta}} [(x_{t-j/3} - \gamma_2 F_{t-j/3})^2 e^{-h_{t-j/3}}|\underline{Y}_T] \right. \\ &\quad + E_{\tilde{\theta}} [(F_{t-j/3} - \rho F_{t-(j+1)/3})^2|\underline{Y}_T] \\ &\quad \left. \left. + \frac{1}{\nu^2} E_{\tilde{\theta}} [(h_{t-j/3} - \rho_{SV} h_{t-(j+1)/3})^2|\underline{Y}_T] \right\} \right), \end{aligned}$$

up to an additive constant. The last equation can be expressed as:

$$\begin{aligned}
Q(\theta|\tilde{\theta}) = & -\frac{1}{2} \left(T \log(1 + d^2 + d^4) + T \log \sigma_y^2 + 3T \log \sigma_x^2 + 3T \log \nu^2 \right. \\
& + \sum_{t=1}^T \sum_{j=0}^2 E_{\tilde{\theta}} [h_{t-j/3} | \underline{Y}_T] \\
& + \frac{1}{(1 + d^2 + d^4) \sigma_y^2} \sum_{t=1}^T \left\{ (y_t - d^3 y_{t-1})^2 \right. \\
& - 2\gamma_1 \left(E_{\tilde{\theta}} [F_t | \underline{Y}_T] y_t - d^3 (E_{\tilde{\theta}} [F_{t-1} | \underline{Y}_T] y_t + E_{\tilde{\theta}} [F_t | \underline{Y}_T] y_{t-1}) + d^6 E_{\tilde{\theta}} [F_{t-1} | \underline{Y}_T] y_{t-1} \right) \\
& + \gamma_1^2 \left(E_{\tilde{\theta}} [F_t^2 | \underline{Y}_T] - 2d^3 E_{\tilde{\theta}} [F_t F_{t-1} | \underline{Y}_T] + d^6 E_{\tilde{\theta}} [F_{t-1}^2 | \underline{Y}_T] \right) \Big\} \\
& + \frac{1}{\sigma_x^2} \sum_{t=1}^T \sum_{j=0}^2 \left\{ x_{t-j/3}^2 E_{\tilde{\theta}} [e^{-h_{t-j/3}} | \underline{Y}_T] - 2\gamma_2 x_{t-j/3} E_{\tilde{\theta}} [F_{t-j/3} e^{-h_{t-j/3}} | \underline{Y}_T] \right. \\
& + \gamma_2^2 E_{\tilde{\theta}} [F_{t-j/3}^2 e^{-h_{t-j/3}} | \underline{Y}_T] \\
& + E_{\tilde{\theta}} [F_{t-j/3}^2 | \underline{Y}_T] - 2\rho E_{\tilde{\theta}} [F_{t-j/3} F_{t-(j+1)/3} | \underline{Y}_T] + \rho^2 E_{\tilde{\theta}} [F_{t-(j+1)/3}^2 | \underline{Y}_T] \\
& + \frac{1}{\nu^2} E_{\tilde{\theta}} [h_{t-j/3}^2 | \underline{Y}_T] - 2\rho_{SV} E_{\tilde{\theta}} [h_{t-j/3} h_{t-(j+1)/3} | \underline{Y}_T] + \rho_{SV}^2 E_{\tilde{\theta}} [h_{t-(j+1)/3}^2 | \underline{Y}_T] \Big\} \Bigg). \tag{B.17}
\end{aligned}$$

From equation (B.17) we note that the estimation step requires the smoothing distribution of the factor path, in order to compute the conditional expectations $E_{\tilde{\theta}} [\cdot | \underline{Y}_T]$. As an exact smoother is not available, in Section B.4 we propose a recursive particle smoother to compute $E_{\tilde{\theta}} [\cdot | \underline{Y}_T]$.

B.3.2 The M-step

By maximizing function $\theta \rightarrow Q(\theta|\tilde{\theta})$, for $\tilde{\theta} = \hat{\theta}^{(i)}$ in equation (B.17), we get the following estimates of the model parameters collected in vector $\hat{\theta}^{(i+1)}$:

$$\begin{aligned}
\hat{\gamma}_2 &= \frac{\sum_{t=1}^T \sum_{j=0}^2 E_{\tilde{\theta}} [F_{t-j/3} e^{-h_{t-j/3}} | \underline{Y}_T] x_{t-j/3}}{\sum_{t=1}^T \sum_{j=0}^2 E_{\tilde{\theta}} [F_{t-j/3}^2 e^{-h_{t-j/3}} | \underline{Y}_T]}, \\
\hat{\sigma}_x^2 &= \frac{1}{3T} \sum_{t=1}^T \sum_{j=0}^2 \left(E_{\tilde{\theta}} [e^{-h_{t-j/3}} | \underline{Y}_T] x_{t-j/3}^2 - 2\hat{\gamma}_2 E_{\tilde{\theta}} [F_{t-j/3} e^{-h_{t-j/3}} | \underline{Y}_T] x_{t-j/3} \right. \\
&\quad \left. + \hat{\gamma}_2^2 E_{\tilde{\theta}} [F_{t-j/3}^2 e^{-h_{t-j/3}} | \underline{Y}_T] \right), \\
\hat{\rho} &= \frac{\sum_{t=1}^T \sum_{j=0}^2 E_{\tilde{\theta}} [F_{t-j/3} F_{t-(j+1)/3} | \underline{Y}_T]}{\sum_{t=1}^T \sum_{j=0}^2 E_{\tilde{\theta}} [F_{t-(j+1)/3}^2 | \underline{Y}_T]}, \\
\hat{\rho}_{SV} &= \frac{\sum_{t=1}^T \sum_{j=0}^2 E_{\tilde{\theta}} [h_{t-j/3} h_{t-(j+1)/3} | \underline{Y}_T]}{\sum_{t=1}^T \sum_{j=0}^2 E_{\tilde{\theta}} [h_{t-(j+1)/3}^2 | \underline{Y}_T]}, \\
\hat{\nu}^2 &= \frac{1}{3T} \sum_{t=1}^T \sum_{j=0}^2 \left(E_{\tilde{\theta}} [h_{t-j/3}^2 | \underline{Y}_T] - 2\hat{\rho}_{SV} E_{\tilde{\theta}} [h_{t-j/3} h_{t-(j+1)/3} | \underline{Y}_T] \right. \\
&\quad \left. + \hat{\rho}_{SV}^2 E_{\tilde{\theta}} [h_{t-(j+1)/3}^2 | \underline{Y}_T] \right),
\end{aligned}$$

and:

$$\begin{aligned}
(\hat{\gamma}_1, \hat{d}, \hat{\sigma}_y) &= \arg \min_{\gamma_1, d, \sigma_y} \left[T \log(1 + d^2 + d^4) + T \log \sigma_y^2 + \frac{1}{(1 + d^2 + d^4) \sigma_y^2} \sum_{t=1}^T \left\{ (y_t - d^3 y_{t-1})^2 \right. \right. \\
&\quad \left. - 2\gamma_1 \left(E_{\tilde{\theta}} [F_t | \underline{Y}_T] y_t - d^3 (E_{\tilde{\theta}} [F_{t-1} | \underline{Y}_T] y_t + E_{\tilde{\theta}} [F_t | \underline{Y}_T] y_{t-1}) + d^6 E_{\tilde{\theta}} [F_{t-1} | \underline{Y}_T] y_{t-1} \right) \right. \\
&\quad \left. \left. + \gamma_1^2 \left(E_{\tilde{\theta}} [F_t^2 | \underline{Y}_T] - 2d^3 E_{\tilde{\theta}} [F_t F_{t-1} | \underline{Y}_T] + d^6 E_{\tilde{\theta}} [F_{t-1}^2 | \underline{Y}_T] \right) \right\} \right]. \tag{B.18}
\end{aligned}$$

The estimates $\hat{\gamma}_2$, $\hat{\sigma}_x^2$, $\hat{\rho}$, $\hat{\rho}_{SV}$, $\hat{\nu}^2$ of the parameters in the M-step are available in closed form, therefore they do not contribute any substantial computational cost. The parameters $\hat{\gamma}_1$, \hat{d} and $\hat{\sigma}_y$, are estimated solving numerically the minimization problem in equation (B.18), with a negligible computational cost compared that of the filtering and smoothing algorithms proposed in the next Section B.4.

B.4 Sequential particle filtering and smoothing algorithms

The E-step in the EM algorithm involves the smoothing distribution of the latent factors paths to compute the conditional expectation $E_{\hat{\theta}}[\cdot | Y_T]$. As an exact smoother is not available for our nonlinear SV model, we propose a sequential backward smoothing algorithm to approximate these conditional expectations. The smoothing algorithm requires, at each date $t - j/3$, for $t = 1, \dots, T$ and $j = 0, 1, 2$, samples from the filtering distribution of the latent factors. For this reason we start with the description of the sequential filtering algorithm based on simulation, before describing the smoothing algorithm. The filtering algorithm proposed in the next section is based on Appendix A.1 in Kim and Stoffer (2008), in particular see their pages 816, 817, 828 and 829, and the references therein, mainly Kitagawa (1996) and Kitagawa and Sato (2001). The idea is to approximate the filtering distribution by a sample of S draws (particles) from it, with S large. This requires an algorithm to draw from the specific distributions of our model. At the E-step of the i -th iteration of the EM algorithm, the estimate of the model parameter $\hat{\theta}^{(i)}$ is available from the previous iteration ($i - 1$)-th.

In this Section, it is convenient to write the model in state space at high frequency. Let $\tau = t - j/3$, for $t = 1, \dots, T$ and $j = 0, 1, 2$. The measurement is $Y_\tau = (y_\tau, x_\tau)'$ if $\tau = t$, and $Y_\tau = x_\tau$ if $\tau = t - j/3$, $j = 1, 2$. The latent factor is $f_\tau = (F_\tau, h_\tau)'$. The transition equation can be written as:

$$f_\tau = \begin{bmatrix} F_\tau \\ h_\tau \end{bmatrix} = \begin{bmatrix} \rho & 0 \\ 0 & \rho_{SV} \end{bmatrix} \begin{bmatrix} F_{\tau-1/3} \\ h_{\tau-1/3} \end{bmatrix} + \begin{bmatrix} \eta_\tau \\ \xi_\tau \end{bmatrix}, \quad \begin{bmatrix} \eta_\tau \\ \xi_\tau \end{bmatrix} \sim i.i.d. \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & \nu^2 \end{bmatrix} \right). \quad (\text{B.19})$$

B.4.1 Sequential filtering based on importance sampling

We propose an algorithm to obtain the samples $f_\tau^{s,(i)} = [F_\tau^{s,(i)}, h_\tau^{s,(i)}]'$, with $s = 1, \dots, S$, from the filtering distribution of the latent factors for parameter value $\hat{\theta}^{(i)}$, denoted as $\ell(f_\tau | Y_\tau; \hat{\theta}^{(i)})$, for any τ . The following steps constitute the filtering algorithm based on importance sampling with resampling:

1. Start at the first date $\tau = t - j/3 = 0$ by drawing a sample $f_0^{s,(i)}$, for $s = 1, \dots, S$, from the stationary distribution of f_τ for parameter value $\hat{\theta}^{(i)}$, denoted $\ell(f_\tau; \hat{\theta}^{(i)})$:

$$\ell(f_\tau; \hat{\theta}^{(i)}) \sim \mathcal{N} \left(0, \begin{bmatrix} \frac{1}{1-\hat{\rho}^{(i),2}} & 0 \\ 0 & \frac{\hat{\nu}^{(i),2}}{1-\hat{\rho}_{SV}^{(i),2}} \end{bmatrix} \right). \quad (\text{B.20})$$

2. At date $\tau = t - j/3 \geq 1/3$, let the input be an approximation of the filtering distribution $\ell(f_{\tau-1/3} | Y_{\tau-1/3}; \hat{\theta}^{(i)})$ via particles $f_{\tau-1/3}^{s,(i)}$, for $s = 1, \dots, S$.

(a) Generate a sample $f_\tau^{0,s,(i)}$, $s = 1, \dots, S$, from $\ell(f_\tau | Y_{\tau-1/3}; \hat{\theta}^{(i)})$. We use $f_\tau | f_{\tau-1/3}, Y_{\tau-1/3} \sim g(\cdot | f_{\tau-1/3})$ where g is the transition density (see Section B.2.2).

Hence, we draw $f_\tau^{0,s,(i)}$ from $g(\cdot | f_{\tau-1/3}^{s,(i)})$. This is achieved by the following steps:

(a.1)

Generate independent random numbers:

$$\eta_\tau^{s,(i)} \sim \mathcal{N}(0, 1), \quad \xi_\tau^{s,(i)} \sim \mathcal{N}(0, \hat{\nu}^{(i),2}),$$

for $s = 1, \dots, S$.

(a.2)

Compute

$$f_\tau^{0,s,(i)} = \begin{bmatrix} F_\tau^{0,s,(i)} \\ h_\tau^{0,s,(i)} \end{bmatrix} = \begin{bmatrix} \hat{\rho}^{(i)} & 0 \\ 0 & \hat{\rho}_{SV}^{(i)} \end{bmatrix} \begin{bmatrix} F_{\tau-1/3}^{s,(i)} \\ h_{\tau-1/3}^{s,(i)} \end{bmatrix} + \begin{bmatrix} \eta_\tau^{s,(i)} \\ \xi_\tau^{s,(i)} \end{bmatrix},$$

for $s = 1, \dots, S$.

(b) Generate a sample from the filtering distribution $\ell(f_\tau | Y_\tau; \hat{\theta}^{(i)})$. We use $\ell(f_\tau | Y_\tau; \hat{\theta}^{(i)}) \propto \ell(Y_\tau | Y_{\tau-1/3}, \underline{f}_\tau; \hat{\theta}^{(i)}) \ell(f_\tau | Y_{\tau-1/3}; \hat{\theta}^{(i)})$ and the importance sampling principle. Compute the weights:

$$\begin{aligned} w_\tau^{s,(i)} &\propto l(Y_\tau | Y_{\tau-1/3}, \underline{f}_\tau^{0,s,(i)}; \hat{\theta}^{(i)}) \\ &= \begin{cases} \frac{1}{\sqrt{(2\pi)^2 \hat{\sigma}_y^{(i),2} \hat{\sigma}_x^{(i),2} \exp\{h_t^{0,s,(i)}\}}} \\ \times \exp \left\{ -\frac{[y_t - d^3 y_{t-1} - \hat{\gamma}_1^{(i)} (F_t^{0,s,(i)} - d^3 F_{t-1}^{0,s,(i)})]^2}{2\hat{\sigma}_y^{(i),2}} - \frac{(x_t - \hat{\gamma}_2^{(i)} F_t^{0,s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\{h_t^{0,s,(i)}\}} \right\} & \tau = t, \\ \frac{1}{\sqrt{2\pi \hat{\sigma}_x^{(i),2} \exp\{h_\tau^{0,s,(i)}\}}} \exp \left\{ -\frac{(x_{t-j/3} - \hat{\gamma}_2^{(i)} F_\tau^{0,s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\{h_\tau^{0,s,(i)}\}} \right\} & \tau = t - j/3, \ j = 1, 2, \end{cases} \end{aligned}$$

for $s = 1, \dots, S$.

Then, generate $f_\tau^{s,(i)} = [F_\tau^{s,(i)}, h_\tau^{s,(i)}]'$ by resampling from $f_\tau^{0,s,(i)} = [F_\tau^{0,s,(i)}, h_\tau^{0,s,(i)}]'$ with weights $w_\tau^{s,(i)}$, for $s = 1, \dots, S$.

This filtering algorithm is straightforward to implement for our model because it only requires (i) to simulate from the state transition density and (ii) evaluate the measurement density. In unreported Monte Carlo experiments, we find that the direct application of this filtering algorithm produces, in a non negligible fraction of the MC replications, degenerate filtered distribution of the latent factors. This degeneracy problem has been solved by modifying the algorithm presented in this section as an auxiliary particle filter algorithm, similarly as Pitt and Shephard (1999). See, among others, Douc, Moulines, and Olsson (2009), Carvalho, Johannes, Lopes, and Polson (2010), Doucet (2010), Lopes and Tsay (2011), Creal (2012), Kantas, Doucet, Singh, Maciejowski, and Chopin (2015), and the reference therein, for a more extensive description of auxiliary particle filter. In Section B.4.2 we describe the auxiliary particle filter used to produce the MC results in the main body of this paper.

B.4.2 Sequential filtering based on auxiliary particle filter

The following steps constitute our auxiliary particle filter:

1. Start at the first date $\tau = t - j/3 = 0$ by drawing a sample $f_0^{s,(i)}$, for $s = 1, \dots, S$, from the stationary

distribution of f_τ for parameter value $\hat{\theta}^{(i)}$, denoted $\ell(f_\tau; \hat{\theta}^{(i)})$ and given in (B.20).

2. At date $\tau = t - j/3 \geq 1/3$, let the input be an approximation of the filtering distribution $\ell(f_{\tau-1/3} | \underline{Y_{\tau-1/3}}; \hat{\theta}^{(i)})$ via particles $f_{\tau-1/3}^{s,(i)}$, for $s = 1, \dots, S$.

(a) Generate auxiliary particles $\bar{f}_\tau^{s,(i)} = [\bar{F}_\tau^{s,(i)}, \bar{h}_\tau^{s,(i)}]'$, where $\bar{F}_\tau^{s,(i)} = \hat{\rho}^{(i)} F_{\tau-1/3}^{s,(i)}$ and $\bar{h}_\tau^{(i)} = \hat{\rho}_{SV}^{(i)} h_{\tau-1/3}^{s,(i)}$, i.e. $\bar{F}_\tau^{s,(i)} = E[F_\tau | F_{\tau-1/3} = F_{\tau-1/3}^{s,(i)}; \hat{\theta}^{(i)}]$ and $\bar{h}_\tau^{s,(i)} = E[h_\tau | h_{\tau-1/3} = h_{\tau-1/3}^{s,(i)}; \hat{\theta}^{(i)}]$.

(b) The auxiliary particles are used to define weights and resample from the old particles $f_{\tau-1/3}^{s,(i)}$. Specifically, compute the weights:

$$\begin{aligned} \check{w}_\tau^{s,(i)} &\propto \ell(Y_\tau | \underline{Y_{\tau-1/3}}, \bar{f}_\tau^{s,(i)}; \hat{\theta}^{(i)}) \\ &= \begin{cases} \frac{1}{\sqrt{(2\pi)^2 \hat{\sigma}_y^{(i),2} \hat{\sigma}_x^{(i),2} \exp\{\bar{h}_t^{s,(i)}\}}} \\ \times \exp \left\{ -\frac{[y_t - d^3 y_{t-1} - \hat{\gamma}_1^{(i)} (\bar{F}_t^{s,(i)} - d^3 \bar{F}_{t-1}^{(i)})]^2}{2 \hat{\sigma}_y^{(i),2}} - \frac{(x_t - \hat{\gamma}_2^{(i)} \bar{F}_t^{s,(i)})^2}{2 \hat{\sigma}_2^{(i),2} \exp\{\bar{h}_t^{s,(i)}\}} \right\} & \tau = t, \\ \frac{1}{\sqrt{2\pi \hat{\sigma}_x^{(i),2} \exp\{\bar{h}_\tau^{s,(i)}\}}} \exp \left\{ -\frac{(x_\tau - \hat{\gamma}_2^{(i)} \bar{F}_\tau^{s,(i)})^2}{2 \hat{\sigma}_2^{(i),2} \exp\{\bar{h}_\tau^{s,(i)}\}} \right\} & \tau = t - j/3, j = 1, 2 \end{cases}, \end{aligned}$$

for $s = 1, \dots, S$. Generate particles $\check{f}_{\tau-1/3}^{s,(i)} = [\check{F}_{\tau-1/3}^{s,(i)}, \check{h}_{\tau-1/3}^{s,(i)}]'$ by resampling $f_{\tau-1/3}^{s,(i)} = [F_{\tau-1/3}^{s,(i)}, h_{\tau-1/3}^{s,(i)}]'$ with weights $\check{w}_\tau^{s,(i)}$, $s = 1, \dots, S$.

(c) Generate a sample from $\ell(f_\tau | \underline{Y_{\tau-1/3}}; \hat{\theta}^{(i)})$. We use $f_\tau | f_{\tau-1/3}, \underline{Y_{\tau-1/3}} \sim g(\cdot | f_{\tau-1/3})$. We draw $f_\tau^{0,s,(i)}$ from $g(\cdot | \check{f}_{\tau-1/3}^{s,(i)})$. This is achieved by:

(c.1)

Generate independent random numbers:

$$\eta_\tau^{s,(i)} \sim \mathcal{N}(0, 1), \quad \xi_\tau^{s,(i)} \sim \mathcal{N}(0, \hat{\nu}^{(i),2}),$$

for $s = 1, \dots, S$.

(c.2)

Compute

$$f_\tau^{0,s,(i)} = \begin{bmatrix} F_\tau^{0,s,(i)} \\ h_\tau^{0,s,(i)} \end{bmatrix} = \begin{bmatrix} \hat{\rho}^{(i)} & 0 \\ 0 & \hat{\rho}_{SV}^{(i)} \end{bmatrix} \begin{bmatrix} \check{F}_{\tau-1/3}^{s,(i)} \\ \check{h}_{\tau-1/3}^{s,(i)} \end{bmatrix} + \begin{bmatrix} \eta_\tau^{s,(i)} \\ \xi_\tau^{s,(i)} \end{bmatrix},$$

for $s = 1, \dots, S$.

(d) Compute the weights:

$$\begin{aligned}
w_\tau^{s,(i)} &\propto \frac{\ell\left(Y_\tau | Y_{\tau-1/3}, \underline{f}_\tau^{0,s,(i)}; \hat{\theta}^{(i)}\right)}{\ell\left(Y_\tau | Y_{\tau-1/3}, \underline{\bar{f}}_\tau^{s,(i)}; \hat{\theta}^{(i)}\right)} \\
&= \begin{cases} \frac{\sqrt{\exp\{\bar{h}_\tau^{s,(i)}\}}}{\sqrt{\exp\{h_\tau^{0,s,(i)}\}}} \exp\left\{-\frac{[y_t - d^3 y_{t-1} - \hat{\gamma}_1^{(i)}(F_t^{0,s,(i)} - d^3 F_{t-1}^{0,s,(i)})]^2}{2\hat{\sigma}_y^{(i),2}} - \frac{(x_t - \hat{\gamma}_2^{(i)} F_t^{0,s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\{h_t^{0,s,(i)}\}}\right\} \\ \times \exp\left\{+\frac{[y_t - d^3 y_{t-1} - \hat{\gamma}_1^{(i)}(\bar{F}_t^{s,(i)} - d^3 \bar{F}_{t-1}^{s,(i)})]^2}{2\hat{\sigma}_y^{(i),2}} + \frac{(x_t - \hat{\gamma}_2^{(i)} \bar{F}_t^{s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\{\bar{h}_t^{s,(i)}\}}\right\} & \tau = t, \\ \frac{\sqrt{\exp\{\bar{h}_{\tau-1/3}^{s,(i)}\}}}{\sqrt{\exp\{h_\tau^{0,s,(i)}\}}} \\ \times \exp\left\{-\frac{(x_\tau - \hat{\gamma}_2^{(i)} F_\tau^{0,s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\{h_\tau^{0,s,(i)}\}} + \frac{(x_\tau - \hat{\gamma}_2^{(i)} \bar{F}_\tau^{s,(i)})^2}{2\hat{\sigma}_2^{(i),2} \exp\{\bar{h}_\tau^{s,(i)}\}}\right\} & \tau = t - j/3, j = 1, 2, \end{cases}
\end{aligned}$$

Generate $f_\tau^{s,(i)} = [F_\tau^{s,(i)}, h_\tau^{s,(i)}]'$ by resampling from $f_\tau^{0,s,(i)} = [F_\tau^{0,s,(i)}, h_\tau^{0,s,(i)}]'$ with weights $w_\tau^{s,(i)}$, for $s = 1, \dots, S$.

B.4.3 Sequential smoothing with importance sampling

Any EM algorithm requires the computation of the smoothing distribution of the latent factor path only once at each iteration i . In our specific case, we need to compute only some moments of the smoothing distribution of the latent factor path. Specifically we need to sample from the smoothing distribution of the factors at each high frequency date $\tau = t - j/3$, with $t = 1, \dots, T$, and $j = 0, 1, 2$. Let $\tilde{f}_\tau^{s,(i)} = (\tilde{F}_\tau^{s,(i)}, \tilde{h}_\tau^{s,(i)})'$, with $s = 1, \dots, S$, be a sample from the smoothing distribution of the latent factors at time τ , obtained using the estimated model parameters $\hat{\theta}^{(i)}$, i.e. during the i -th iteration of the EM algorithm. Then, the conditional expectations $E_{\hat{\theta}}[\cdot | Y_T]$ of the E-step for $\hat{\theta} = \hat{\theta}^{(i)}$ can be approximated by sample averages of the S particles.

The smoothing algorithm is based on Appendix A.2 in Kim and Stoffer (2008) and the reference therein, mainly Godsill, Doucet, and West (2004), and is free from degeneracy. It uses the particles of the filtering distribution as input. Specifically, let $f_\tau^{s,(i)} = (F_\tau^{s,(i)}, h_\tau^{s,(i)})'$, with $s = 1, \dots, S$, be draws from the filtering distribution $\ell(f_\tau | Y_\tau; \hat{\theta}^{(i)})$ for date $\tau = t - j/3$, $j = 0, 1, 2$ (see Section B.4.2). The following steps constitute the backward sequential smoothing algorithm. For any $s = 1, \dots, S$:

1. Start at the last date $\tau = T$, and draw $\tilde{f}_T^{s,(i)}$ from set $\{f_T^{r,(i)}, r = 1, \dots, S\}$ with equal weights $1/S$. In other words, we obtain one draw from the filtering distribution $\ell(f_T | Y_T; \hat{\theta}^{(i)})$ at the last sample date.
2. For any date τ , from $\tau = T - 1/3$ to $\tau = 0$:

(a) Compute the weights:

$$w_{\tau, \tau+1/3}^{r,(i)} \propto g(\tilde{f}_{\tau+1/3}^{s,(i)} | f_\tau^{r,(i)}; \hat{\theta}^{(i)}),$$

for $r = 1, \dots, S$.

(b) Draw $\tilde{f}_\tau^{s,(i)}$ from $\{f_\tau^{r,(i)}, r = 1, \dots, S\}$ with probability weights $\{w_{\tau, \tau+1/3}^{r,(i)}, r = 1, \dots, S\}$.

At the end of the smoothing algorithm we have S simulated paths $\{(\tilde{f}_0^{s,(i)}, \tilde{f}_{1/3}^{s,(i)}, \dots, \tilde{f}_T^{s,(i)}), s = 1, \dots, S\}$ from the smoothing distribution $\ell(f_0, f_{1/3}, \dots, f_T | \underline{Y_T}; \hat{\theta}^{(i)})$. Note that the second step of our backward sequential smoothing algorithm requires only (i) a sample from the filtering distribution which is already available from the filtering algorithm, and (ii) to be able to evaluate the transition density.

B.4.4 Stopping rule

For the Monte Carlo EM algorithm to converge to the MLE estimate, the number of particles S needs to increase with the number of EM iterations, see for instance Olsson, Cappé, Douc, and Moulines (2008), Neath (2013) and the references therein. Moreover, a rule needs to be set in order to stop the algorithm and assess its convergence. We follow the same procedure of Kim and Stoffer (2008). On the basis of the work of Chan and Ledolter (1995), Kim and Stoffer (2008) start the EM algorithm with a small value of S to save computing time, at the end of each EM iteration compute ϵ - the estimated change in log-likelihood with respect to the previous EM iteration - and increase S when ϵ is below a certain small lower bound. The EM algorithm in our paper is implemented starting with a number of particles $S = 500$, then the value of S is increased to 1000 as soon as $\epsilon < 0.10$ for a certain iteration, then S is increased to 1500 when $\epsilon < 0.05$, and finally the algorithm is stopped at the first iteration in which $\epsilon < 0.01$. The values of ϵ and S , together with the stopping rule, were calibrated in preliminary unreported MC experiments. See Kim and Stoffer (2008) and Chan and Ledolter (1995) for an in-depth analysis of this procedure and the concept of “Relative Likelihood”.

Appendix C: Mixed frequency linear state space model with one flow sampled low frequency variable

In this Appendix we show one way to adapt the measurement and transition equations of the linear state space model with mixed frequency data in Bai, Ghysels, and Wright (2013), for the case of one low frequency variable which is flow sampled. This adaptation is necessary to implement the Kalman filter algorithm for ML estimation of the structural model without SV used in the empirical application of Section 6.

C.1 Model setup

The linear state space model without SV considered in the empirical application in Section 6 has one flow sampled low frequency variable y_t , with $t = 1, \dots, T$, $m = 3$ high frequency subperiods and a single latent factor (i.e. $n_f = 1$). The latent factor follows an $AR(1)$ process:

$$F_{t+j/3} = \Phi_1 F_{t+(j-1)/3} + \eta_{t+j/3} \quad t = 1, \dots, T, \quad j = 0, 1, 2, \quad (\text{C.1})$$

where Φ_1 is a scalar parameter to be estimated. Latent process y^* is related to the factor as follows:

$$y_{t+j/3}^* = \gamma_1 F_{t+j/3} + u_{1,t+j/3} \quad t = 1, \dots, T, \quad j = 0, 1, 2, \quad (\text{C.2})$$

with $u_{1,t+j/3}$ having an $AR(1)$ representation:

$$u_{1,t+j/3} = d_1 u_{1,t+(j-1)/3} + \varepsilon_{1,t+j/3}. \quad (\text{C.3})$$

The observed low-frequency process y is flow sampled, i.e. it relates to the latent process y^* in the following way:

$$y_t = y_t^* + y_{t-1/3}^* + y_{t-2/3}^*, \quad t = 1, \dots, T, \quad j = 0, 1, 2. \quad (\text{C.4})$$

The high frequency process $x_{t+j/3}$ relates to the factor as follows:

$$x_{t+j/3} = \gamma_2 F_{t+j/3} + u_{2,t+j/3} \quad t = 1, \dots, T, \quad j = 0, 1, 2, \quad (\text{C.5})$$

where:

$$u_{2,t+j/3} = d_2 u_{2,t+(j-1)/3} + \varepsilon_{2,t+j/3}. \quad (\text{C.6})$$

The innovations (η) , (ε_1) , (ε_2) are mutually independent i.i.d. Gaussian processes, with distributions $\mathcal{N}(0, 1)$, $\mathcal{N}(0, \sigma_{\varepsilon_1}^2)$ and $\mathcal{N}(0, \sigma_{\varepsilon_2}^2)$.

C.2 State space representation and Kalman filter

The above equations yield a periodic state space model with measurement equation:

$$Y_t^j = Z_j \alpha_{t+j/m} \quad \begin{cases} Y_t^j = (y_t, x_t)' & j = 0 \\ Y_t^j = x_{t+j/m} & j = 1, 2 \end{cases}, \quad (\text{C.7})$$

for $t = 1, \dots, T$, where

$$Z_0 = \begin{bmatrix} \gamma_1 & \gamma_1 & \gamma_1 & 1 & 1 & 1 & 0 \\ \gamma_2 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$Z_j = \begin{bmatrix} \gamma_2 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad j = 1, 2,$$

and state vector:

$$\alpha_{t+j/3} = (F_{t+j/3}, F_{t+(j-1)/3}, F_{t+(j-2)/3}, u_{1,t+j/3}, u_{1,t+(j-1)/3}, u_{1,t+(j-2)/3}, u_{2,t+j/3})'.$$

The transition equation is:

$$\alpha_{t+j/m} = R\alpha_{t+(j-1)/3} + Q\zeta_{t+j/3}, \quad t = 1, \dots, T, \quad j = 0, 1, 2, \quad (\text{C.8})$$

where

$$R = \begin{bmatrix} \Phi_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & d_2 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$\zeta_{t+j/m} = (\eta_{t+j/m}, \varepsilon_{1,t+j/m}, \varepsilon_{2,t+j/m})'$, and $\Sigma_\zeta = \text{diag}(1, \sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2)$ denotes the variance-covariance matrix of $\zeta_{t+j/m}$. Then, the Kalman filter algorithm presented in Appendix A can be performed after replacing matrices Z_0 , Z_j , R , Q and Σ_ζ by the new definitions.